Estimating Foreign-Born Emigration from the United States Using Data from the American Community Survey

Mark A. Leach
Eric B. Jensen

Population Division
U.S. Census Bureau
4600 Silver Hill Road
Washington, D.C. 20233
mark.a.leach@census.gov
eric.b.jensen@census.gov

Paper presented at the Federal Committee on Statistical Methodology Research Conference
November 4-6, 2013
Washington, D.C.

Abstract

In the absence of comprehensive data on foreign-born persons who leave the United States, demographers often use a residual method to estimate annual rates of foreign-born emigration. Historically, demographers used data from two subsequent decennial censuses to enumerate the foreign-born population in an earlier census, survive that population forward ten years, and compare the expected survived population to the enumerated population in the second Census. The difference, or residual, between the survived and enumerated populations, excluding immigrants who arrived in the interim, is assumed to be due to emigration. The main critique of this method is that only a small proportion of emigration that occurs within ten years of first arriving in the United States, common among groups such as the Mexican born, is measured. Annual data from the American Community Survey, however, may provide sufficient sample size of the foreign-born to overcome this weakness of the residual method. The present study assesses the feasibility of using data from the American Community Survey to estimate foreign-born emigration and presents comparisons with estimates based on other data sources and methods.

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Introduction

One of the more difficult tasks that demographers must undertake to estimate change in a national population is to measure the number of people that leave, or emigrate, to reside in another country. In the United States, immigration is a large driver of emigration. In other words, some immigrants who come to the United States stay only temporarily, as in the cases of university students or those with temporary work visas. Other immigrants who come here to work for twenty or thirty years may eventually return home after retirement. However long the duration of their stay and whatever the reasons for leaving, the number of people that emigrate is one of the most difficult components of population change to measure. The U.S. government does not track systematically people who leave the country, and the number of places to which they go makes individual data collection impossible.

The purpose of this research is to assess the feasibility of using data from the American Community Survey (ACS) to estimate foreign-born emigration from the United States. In the absence of comprehensive data on foreign-born persons who leave the United States, demographers have used the residual method to measure indirectly foreign-born emigration. Data from two subsequent decennial censuses are used typically to compare a foreign-born population enumerated at two points in time after accounting for mortality and new immigrant arrivals. The difference, or residual, is assumed to be due to emigration. While theoretical weaknesses of the residual method have long been known, publication in recent years of emigration estimates based on other data sources and methods have shed further light on potential bias in residual-based emigration estimates.

This research is part of work at the U.S. Census Bureau to improve its method of estimating foreign-born emigration for its Population Estimates Program. For its Vintage 2012 population estimates, the Census Bureau used data from Census 2000 and several years of ACS microdata to estimate emigration levels and rates. Continuing to rely on Census 2000, however, presents similar problems as decennial-to-decennial estimates. We use annual data from the ACS to estimate foreign-born emigration rates. Emigration rates based on annual ACS data overcome weaknesses of the residual method by measuring emigration over a shorter period of time. This results in more timely emigration estimates by including immigrants who arrived in the United States relatively recently (Van Hook et al. 2006). A shorter observation period also better reflects the typical timing of emigration relative to immigrants' arrival in the United States. Return migration often occurs within the first few years after arrival in the United States, especially among Mexican immigrants who are relatively more likely than other immigrants to return home (Massey, Durand and Malone 2002; Riosmena 2004). By averaging emigration over ten years, however, a decennial-to-decennial estimate implicitly assumes that emigration is distributed evenly over a decade, which likely results in an estimate that is biased downward. Our research findings suggest that reducing the number of years over which a foreign-born population is observed improves residual-based measures of emigration relative to previous residual-based estimates and estimates based on other methods and data sources.

Previous Research and Methods

The residual method was first developed by Warren and Peck (1980) to estimate foreign-born emigration from the United States between 1960 and 1970. They estimated foreign-born emigration by subtracting the observed foreign-born population in the 1970 census from the expected foreign-born population in 1970 derived from the 1960 census. The difference, or residual, is assumed to be due to emigration. The expected population was calculated by subtracting an estimated number of deaths experienced in the foreign-born population between 1960 and 1970 and adding an estimate of new arrivals during the decade. Warren and Peck also adjusted for what the authors referred to as "nativity bias" or when foreign-born respondents misreport their nativity status. They also calculated annual rates of emigration by dividing a residual by the foreign-born population at risk of emigrating, enumerated in the most recent census. Using the 1980 and 1990 censuses, Ahmed and Robinson (1994) refined the method by adjusting assumptions about emigration of foreign born who arrived during the period between 1980 and 1990. In another widely-cited study, Mulder (2003) used the 1990 and 2000 censuses to assess trends in emigration during the 1990s. The U.S. Census Bureau currently uses Census 2000 data in combination with annual data from the ACS to estimate emigration rates for several foreign-born subpopulations (Bhaskar, Arenas-Germosen and Dick 2013; Bhaskar, Rastogi and Kennedy-Puthoff 2008).

There are several common criticisms of the residual method, some of which focus on its reliance on data from decennial censuses. First, in general the residual method is sensitive to coverage or measurement error in census or survey data. For instance, if the coverage of the foreign-born population is relatively poor in an earlier census, the

foreign-born population will be underestimated and an emigration residual will be lower than it otherwise would have been had there been no undercount. This will result in an underestimate of emigration.

Second, a decennial-to-decennial estimate of emigration lacks timeliness both in terms of annual variation on international migration flows and in terms of the cohort of immigrants included in a measure emigration. International migration fluctuates annually due to economic, social and political dynamics in sending and receiving countries (Bean and Stevens 2003). For example, the enumerated Mexican-born population was 9.6 million in Census 2000, grew to an estimated 12.6 million in 2007, and declined to 12.3 million in 2010 (Passel, Cohn and Gonzalez-Barrera 2012), suggesting more emigration occurred at the end of the decade relative to the first half of the decade. A decennial-to-decennial measure of emigration will not reflect this variation (Passel, Cohn and Gonzalez-Barrera 2012; Rendall, Brownell and Kups 2011).

In addition, decennial-to-decennial emigration estimates do not include the most recently-arrived immigrants (Van Hook and Zhang 2011; Van Hook et al. 2006). Estimates produced using 1990 and 2000 census data were based on the foreign-born population that had arrived in the United States prior to 1990. As such, estimates produced in the early 2000s did not include immigrants that arrived since 1990. Previously, researchers either assumed that recent arrivals have similar rates of emigration as earlier arrivals or made assumptions about differences in emigration behavior, neither of which can be validated. In either case, emigration trends among recent arrivals, perhaps the most important group to measure emigration given their propensity to return to their home country, are not captured in the estimate.

Also related to recent arrivals, a criticism of the residual method is that measuring emigration over ten years may not accurately reflect actual patterns of international migration. Return migration typically occurs within just a few years after arriving in the United States, especially in the Mexican-born population which has relatively high rates of return migration (Massey, Durand and Malone 2002; Massey and Singer 1995). The residual method assumes that emigration is evenly distributed across an observation horizon by annualizing total emigration in a period to calculate an annual emigration rate. If most emigration occurs in the first few years of an intercensal period and subsequently declines, average annual emigration will be underestimated by including years for which emigration is minimal. For example, an emigration residual based on 1990 and 2000 Censuses measures emigration behavior of the foreign-born population that arrived prior to 1990. Most emigration for this cohort of immigrants likely occurred in the beginning of the 1990s when a greater proportion of the cohort had relatively less experience in the United States. To the extent that emigration declined between 1995 and 2000, a 1990-2000 ten-year residual would be only slightly larger than a 1990-1995 five-year residual. When each is annualized, the five-year average will be close to twice that of the ten-year average, and the ten-year average will underestimate emigration.

In response to these weaknesses, researchers have increasingly sought alternative methods and data to measure foreign-born emigration, using administrative data, household surveys, and census data from other countries (Jasso and Rosenzweig 1982; Passel, Cohn and Gonzalez-Barrera 2012; Rendall, Brownell and Kups 2011; Schwabish 2011; Van Hook et al. 2006; Woodrow-Lafield 1996). Jasso and Rosenzweig (1982) used administrative data from the former Immigration and Naturalization Service, now part of the U.S. Department of Homeland Security (DHS), for a cohort (1971) of legal immigrants to the United States. By linking administrative data records in 1979 to members of the 1971 cohort, the authors were able to estimate cumulative emigration rates for that cohort based on attrition from the administrative data as well as estimates of mortality. However, the analysis only included legal immigrants and used data that are not readily available. Woodrow-Lafield (1996) used household surveys that asked respondents to report on the residence status of household members and other close relatives living abroad to estimate emigration from the United States. To accurately derive estimates from surveys with this design (network sampling), a multiplicity adjustment must be calculated to ensure that multiple survey respondents are not reporting the same individual living abroad. In this study, the multiplicity adjustment reduced the initial number of emigrants reported through the household surveys by nearly 80 percent, making the estimates extremely sensitive to the assumptions used to make the adjustment. Massey and Singer (1995) analyzed survey data collected in Mexico from 1987-1992 containing life histories of return migrants. However, it is difficult to compare their method to other methods because it measures gross migration rates (number of trips), not net migration rates (number of people).

More recently, Van Hook et al. (2006) used matched files from the Current Population Survey (CPS) to estimate emigration of the foreign born. The CPS has a quasi-longitudinal design in which the same household is included in the survey for four consecutive months and then rotates out of the survey for eight months; they are then brought

back into the sample for the same four months the following year. The sampling frame for the CPS is made up of addresses, not individuals, so respondents who move to a new address drop out of the CPS sample. Van Hook et al. estimate the probability that a foreign-born household was not followed-up in the subsequent CPS sample because of emigration. Because the CPS contains detailed social and economic data, the method can produce emigration rates by demographic and social characteristics. In addition, the method incorporates the most recent arrivals which is evident in much higher estimated emigration rates relative to rates produced using the residual method.

Schwabish (2011) estimated the probability of emigrating using longitudinal administrative earnings data from the U.S. Social Security Administration (SSA). The method tracks the sequence of earnings over time and identifies periods of positive earnings followed by a period of no earnings as an emigration event. A limitation of the method is that the sample is limited to workers that are part of the Social Security System, which systematically excludes immigrants that are not in the labor force or undocumented immigrants who do not participate in the formal economy.

Researchers also have used survey and census data from other countries, particularly Mexico, to estimate foreign-born emigration from the United States. Rendall, Brownell and Kups (2011) analyzed micro data from the National Survey of Occupation and Employment (ENOE) in Mexico to measure return migration to Mexico during the 2008-2009 economic recession. The ENOE is a quarterly household employment survey where a household remains in the sample for five consecutive quarterly interviews. By analyzing changes to household rosters between quarters, Rendall, Brownell and Kups (2011) were able to measure return migration to Mexico from the United States. Similarly, Passel, Cohn and Gonzalez-Barrera (2012) used data from Mexico's 2010 census which reported a respondent's residence five years ago to show that return migration to Mexico had increased substantially in the late 2000s. We replicate a summary of previous estimates provided by Van Hook et al. (2006) and add more recent estimates in Table 1.

Table 1. Previously Published Rates of Foreign-Born Emigration, 1960 – 2010.

Population and Reference of Estimate	Period	Annual Emigration Rate (%)
All Foreign-Born		
Census-to-Census Residual		
Warren and Peck (1980)	1960 - 1970	1.2
Ahmed and Robinson (1994)	1980 - 1990	1.2
Mulder (2003)	1990 - 2000	0.9
Other Methods		
Van Hook et al. (2006), CPS Matching	1998 - 2004	2.9
Schwabish (2011)	1978 - 1998	1.3 [a]
In U.S. 0 - 10 Years (recent arrivals)		
Census-to-Census Residual		
Warren and Peck (1980)	1960 - 1970	4.4
Mulder (2003)	1990 - 2000	0.3
Other Methods		
Borjas and Bratsberg (1996), Census/INS residual	1970 - 1980	3.2
Van Hook et al. (2006), CPS Matching	1998 - 2004	4.3
Schwabish (2011)	1978 - 1998	2.3 [a]
Mexican Foreign-Born		
Census-to-Census Residual		
Ahmed and Robinson (1994)	1980 - 1990	0.6
Mulder (2003)	1990 - 2000	0.4
Other Methods		
Massey and Singer (1995), Life Histories	1965 - 1989	51.5
Passel, Cohn and Gonzales-Barrera (2012), Mexico 2000 Census	1995 - 2000	0.8 [b]
Van Hook et al. (2006), CPS Matching	1998 - 2004	4.3
Rendall, Brownell and Kups (2011), Mexico ENOE Survey	2005 - 2009	3.7 [c]
Passel, Cohn and Gonzales-Barrera (2012), Mexico 2010 Census	2005 - 2010	1.5 [d]
Schwabish (2011)	1978 - 1998	2.5 [a]
In U.S. 0 - 10 Years (recent arrivals)		
Census-to-Census Residual		
Mulder (2003)	1990 - 2000	0.1
Other Methods		
Massey, Durand and Malone (2002), Event history (Unauthorized)	1965 - 1985	28.8
Massey, Durand and Malone (2002), Event history (Legal)	1965 - 1985	12.5

[a] Average of Schwabish's estimated annual rates for 1978-1998

[b] Rate calculated by annualizing Passel, Cohn and Gonzales-Barrera's estimated total emigration for 1995-2000 and dividing by an estimate of the population at risk of emigrating: the sum of the Mexican-born population arrived prior to 1995, obtained from the 2000 U.S. Census, and Passel, Cohn and Gonzales-Barrera's estimated total emigration for 1995-2000.

[c] Rate calculated by averaging Rendall, Brownell and Kups' (2011) four annual total estimates and dividing by an estimate of the population at risk of emigrating: the estimated Mexican-born population from the 2005 American Community Survey.

[d] Rate calculated by annualizing Passel, Cohn and Gonzales-Barrera's estimated total emigration for 2005-2010 and dividing by an estimate of the population at risk of emigrating: the sum of the Mexican-born population arrived prior to 2005, obtained from the 2005 American Community Survey, and Passel, Cohn and Gonzales-Barrera's estimated total emigration for 2005-2010.

Residual Method of Estimating Emigration

We estimate foreign-born emigration using a residual method similar to that developed by Warren and Peck (1980) and refined by Ahmed and Robinson (1994) and Mulder (2003). The basic equation to estimate emigration between two points in time is

$$E_{t1\text{-}t2} = (P_{t1} - D_{t1\text{-}t2}) - P_{t2}, \quad (1)$$

where $E_{t1\text{-}t2}$ is an emigration residual, or estimated number of foreign-born who emigrated between time 1 and time 2 (we refer to this period as the residual survival period),
P_{t1} is the estimated foreign-born population at time 1,
$D_{t1\text{-}t2}$ is the estimated number of deaths experienced between time 1 and time 2 in the foreign-born population estimated at time 1, such that
P_{t1} - $D_{t1\text{-}t2}$ is the expected survived foreign-born population at time 2 assuming no emigration, and
P_{t2} is the estimated foreign-born population at time 2 that arrived prior to time 1.

Annual rates of emigration are useful to compare estimates based on different methods, data sources, and time periods. Ahmed and Robinson (1994) estimated annual emigration rates for the 1980s by dividing $E_{1980\text{-}1990}$ by 10 to annualize the residual and dividing again by P_{1980}, the estimated at-risk population. In the 2012 Vintage of Population Estimates, the Census Bureau divided an annualized emigration residual by an estimate of the foreign-born population at the mid-point of a residual period. This divisor takes into account deaths and emigration during the residual period, each reducing the population at risk of emigrating. We calculate an annual emigration rate with the following equation:

$$R_{t1\text{-}t2} = E_{t1\text{-}t2} / (PY_{t1\text{-}t2} - 0.5^* E_{t1\text{-}t2})^* 100 \quad (2)$$

where $R_{t1\text{-}t2}$ is an annual rate of emigration between time 1 and time 2, expressed as a percent or number of emigrants per 100 population,
$E_{t1\text{-}t2}$ is an emigration residual calculated using Equation (1), and
$PY_{t1\text{-}t2}$ is total person-years survived in P_{t1} between time 1 and time 2, which accounts for mortality in the at-risk population. Subtracting one half of the residual from person-years survived accounts for emigration during a residual period, which also reduces the population at risk of emigrating at a given point in time.

Estimating Emigration Using Data from the American Community Survey

We use individual-level micro data from the ACS to estimate both P_{t1} and P_{t2} in Equation 1 above. In general, the relatively large sample of the foreign-born population in the ACS offers an advantage over other survey data because foreign-born emigration is a relatively rare event. The residual method in particular requires a large sample because one must estimate population cells by single year of age and by sex to account for mortality between time 1 and time 2. Census data offered a similar advantage.

The primary purpose of this research, however, is to assess whether estimating P_{t1} and P_{t2} using the ACS improves a residual estimate of emigration relative to a decennial-to-decennial estimate. Annual ACS data allow us to calculate emigration rates based on observation periods much shorter than ten years. In fact, one may estimate P_{t1} and P_{t2} using any combination of available ACS micro data, 2005 through 2012. For example, one may estimate P_{t1} using the 2006 ACS and estimate P_{t2} with the 2008 ACS to calculate a 2-year residual. Or one may use the 2008 and 2012 ACS to create a 4-year residual.

To reduce the number of possible combinations, we focus on the 2006 to 2010 ACS samples. The 2006-2010 timeframe serves two primary purposes. First, we believe a five-year observation horizon more closely aligns with previous research that shows emigration is most likely within a few years after arrival in the United States. We expect that annual emigration rates based on one- to four-year observation horizons will lead to better estimates that are more aligned with estimates based on other data and methods. Second, the 2006 to 2010 period in particular allows us to assess the effect of the length of an observation horizon by making comparisons with estimates based

on Census 2000 and the 2010 ACS, which have a ten year observation horizon similar to previous decennial-to-decennial estimates.

We extract household records (which excludes the relatively small population of foreign born that resides in group quarters) from the 2006-2010 five-year ACS micro data file for consistency in population controls and weighting methods and disaggregate the data by survey year.[1] With the five years of sample data, we calculate a series of emigration residuals and annual emigration rates based on two-, three-, and four-year observation horizons.[2] There is only one way to calculate a 4-year residual, using the 2006 and 2010 samples, but there are multiple ways to calculate two- and three-year residuals. For example, one may calculate three two-year residuals based on estimates of P_{t1} in 2006, 2007, and 2008 paired with P_{t2} estimates in 2008, 2009, and 2010, respectively.

Table 2 shows sample sizes, population estimates, and margins of error for estimates of P_{t1} and P_{t2} that we use to calculate emigration residuals using the Census Bureau's internal 2006-2010 ACS micro data file. The table includes samples and estimates by four foreign-born subpopulations for which the Census Bureau estimated foreign-born emigration separately in the Vintage 2012 population estimates: recently-arrived Mexican-born, earlier-arrived Mexican-born, recently-arrived non-Mexican-born, and not recently-arrived non-Mexican-born. To compare our estimates to previously-published estimates, we also calculate rates for the total foreign-born population, for immigrants in the United States for less than 10 years, and for the Mexican-born population. Recent arrival is defined as having a year of entry less than or equal to ten years prior to P_{t1}. Mexican origin is identified by the place of birth question in the ACS.

We expect two-, three-, and four-year emigration residuals and rates to vary for several reasons. First, residual periods with different numbers of years will be based on different populations estimated at either time 1 or time 2, or both. The compositions of two base populations may differ enough to make one population more or less prone to emigration than another, or historical events may affect one base population more than another. Second, the length of a residual period carries implications for the amount of emigration represented in a residual and the number of years over which the migration is averaged. One or both of these factors may affect differences in emigration rates.

Sampling and non-sampling error may affect variability in the rates we calculate. We show margins of error for P_{t1} and P_{t2} population estimates in Table 2. Calculating a true margin of error for a residual-based emigration rate is more difficult, however, given that we use population estimates from multiple samples. To simplify, we construct a pseudo-margin of error for an emigration residual based on the 90-percent confidence intervals of P_{t1} and P_{t2}. First, we calculate an average survival rate for a foreign-born subpopulation and apply that rate to the limits of the confidence interval for P_{t1}. Second, the lowest possible residual estimate between P_{t1} and P_{t2} is the difference between the lower limit of the P_{t1} confidence interval (after applying the survival rate) and the upper limit of the P_{t2} interval. The largest possible residual estimate is the upper limit of P_{t1} and the lower limit of P_{t2}. Third, we divide each residual difference by the denominator in Equation 2 to calculate upper and lower limits for an emigration rate. The pseudo-margin of error of a rate is the absolute difference between the upper (or lower) limit of a rate and R_{t1-t2}.

For example, the estimated size of the recent Mexican-born population in 2007 is 4,771,300 (P_{t1}) with a margin of error of 65,900 (Table 2). After surviving this population forward two years to 2009 by age and sex, we may find that the population as a whole has an average survival rate of 0.98. We then multiply the upper and lower confidence limits of the P_{t1} population estimate, 4,837,200 and 4,705,400 respectively, by the average survival rate of 0.98. For the lower limit of a 2007-2009 two-year residual, we then subtract the upper limit of P_{t2} (4,426,900+62,600, Table 2) from 4,611,292, the product of 4,705,400 and 0.98. The upper limit of the residual is the survived upper limit of P_{t1}, 4,837,200*0.98, minus the lower limit of P_{t2}, 4,426,900-62,600. We then divide each limit for the residual by the

[1] Person weights in a single-year ACS file are controlled to the vintage of population estimates of the same year. Population estimates, and thus weighting controls, for a particular year may change from vintage to vintage due to updated data and change in methods. The Census Bureau does not recalculate person weights for prior single-year ACS files. It does, however, use consistent population controls when it creates each five-year ACS file. For annual population estimates from a five-year ACS file, we multiply each person weight by five.

[2] After a preliminary analysis we decided to forego presenting estimates based on a observation horizon of just one year. We found most of the one-year residuals that we calculated to be negative, which implies that there were more immigrants estimated at time 2 after subtracting deaths from the time 1 population estimate, which is not demographically possible.

denominator in Equation 2 (not shown in a table) to calculate the upper and lower limits for the emigration rate.

In the results section that follows, we first assess differences between 2-year, 3-year, and 4-year emigration rates we estimate using the 2006-2010 ACS. We then compare our ACS-to-ACS estimates to previously-published emigration rates based on other data sources and estimation methods. Finally, we assess the implications of the length of a residual period specifically by comparing our estimates to rates based on Census-to-ACS residuals estimated by the Census Bureau for its Vintage 2012 Population Estimates.

Table 2. Sample Size and Population Estimates for P_{t1} and P_{t2} Used to Estimate Foreign-Born Emigration Rates, United States, 2006 – 2010.

	Base Population (P_{t1})			Enumerated Population in Residual Year (P_{t2})											
				t2 = 2007			t2 = 2008			t2 = 2009			t2 = 2010		
Year (t1)	N	Estimate	MOE[1]	N	Estimate	MOE[1]	N	Estimate	MOE[1]	N	Estimate	MOE[1]	N	Estimate	MOE[1]
Non-Mexican, >10 Years in US															
2006	214,600	16,108,400	84,300	207,100	15,808,400	91,700	202,500	15,591,600	84,600	199,500	15,436,800	78,800	197,600	15,328,700	70,600
2007	215,300	16,481,600	92,700				210,900	16,306,200	86,500	207,400	16,096,700	79,000	206,000	16,036,800	73,200
2008	219,400	17,010,900	88,700							215,900	16,807,600	80,000	214,300	16,726,100	74,200
2009	225,200	17,597,900	80,900										223,800	17,538,100	81,000
Non-Mexican, <= 10 Years in US															
2006	103,700	8,953,100	79,400	98,700	8,594,500	76,700	96,700	8,528,800	83,500	96,900	8,548,400	77,500	99,000	8,701,800	91,200
2007	101,200	8,891,300	84,200				97,900	8,705,200	86,200	98,500	8,771,400	79,600	100,300	8,894,500	94,900
2008	100,000	8,976,800	88,900							99,500	8,935,200	86,300	102,200	9,142,000	96,900
2009	100,000	9,049,400	88,200										102,300	9,210,200	94,600
Mexican, >10 Years in US															
2006	73,400	6,260,200	54,500	71,400	6,222,800	61,300	67,900	5,962,900	51,900	67,800	6,021,300	55,400	67,500	6,066,000	67,700
2007	75,000	6,564,800	62,400				71,400	6,303,600	55,100	71,200	6,344,800	57,200	70,900	6,394,500	68,800
2008	74,500	6,607,200	55,400							74,500	6,671,900	59,600	74,100	6,709,900	69,300
2009	78,600	7,082,800	63,600										78,500	7,152,900	72,900
Mexican, <= 10 Years in US															
2006	48,100	4,855,800	68,500	46,100	4,741,400	64,900	43,200	4,454,100	67,600	43,000	4,413,700	60,900	42,700	4,345,700	69,400
2007	45,800	4,771,300	65,900				42,800	4,449,800	67,300	42,700	4,426,900	62,600	42,200	4,339,100	68,500
2008	42,100	4,416,500	69,900							42,000	4,388,900	65,000	41,800	4,327,300	69,200
2009	39,900	4,207,400	63,400										39,700	4,141,600	71,600

Source: Authors' calculations using the 2006-2010 American Community Survey, unpublished data

Notes: All values shown in table are rounded to the nearest multiple of 100. Unrounded values are used in all calculations. Data based on sample. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see www.census.gov/acs/www/.

[1] Data are based on a sample and are subject to sampling variability. A margin of error is a measure of an estimate's variability. The larger the margin of error is in relation to the size of the estimate, the less reliable the estimate. When added to and subtracted from the estimate, the margin of error forms the 90 percent confidence interval.

Results

Table 3 shows our estimated rates of emigration, expressed as emigrants per 100 foreign-born population, for the total foreign-born population, the foreign-born population by time in the United States, and the foreign-born population by place of birth (Mexican, non-Mexican).

Table 3. Estimated Rates of Foreign-Born Emigration Using the 2006 – 2010 American Community Survey.

	Years in Residual Time Period								
	Two			Three			Four		
Base Population / Year (t1)	t2	Rate[1]	Pseudo-MOE[2]	t2	Rate[1]	Pseudo-MOE[2]	t2	Rate[1]	Pseudo-MOE[2]
All Foreign-Born									
2006	2008	1.62	0.42	2009	0.95	0.27	2010	0.51	0.22
2007	2009	0.78	0.47	2010	0.25	0.33			
2008	2010	-0.57	0.50						
> 10 Years in US									
2006	2008	0.88	0.44	2009	0.40	0.30	2010	0.11	0.22
2007	2009	0.37	0.47	2010	-0.07	0.31			
2008	2010	-0.59	0.44						
<= 10 Years in US									
2006	2008	2.83	0.80	2009	1.85	0.53	2010	1.15	0.43
2007	2009	1.48	0.82	2010	0.81	0.58			
2008	2010	-0.53	0.93						
Non-Mexican									
2006	2008	1.07	0.51	2009	0.61	0.31	2010	0.19	0.25
2007	2009	0.17	0.51	2010	-0.26	0.36			
2008	2010	-0.61	0.54						
Mexican									
2006	2008	2.86	0.79	2009	1.72	0.54	2010	1.24	0.42
2007	2009	2.16	0.89	2010	1.42	0.62			
2008	2010	-0.46	0.89						

Source: Authors' calculations using the 2006-2010 American Community Survey, unpublished data

[1] Emigrants per 100 foreign-born population

[2] It is difficult, if not impossible, to calculate a true margin of error for an emigration rate when using the residual method. We construct a "likely" range of values, which we refer to here as a pseudo-margin of error, based on 90 percent confidence intervals of the population estimates used in the residual calculation.

Rates that are based on population estimates using the 2008 ACS raise our concern. All the rates for which 2008 data are used to estimate P_{t1} are negative, which is not demographically possible. We might assume a negative residual implies zero emigration if zero falls within the margin of error, but this is not the case with most of the negative rates based on 2008 data. Furthermore, the historical context of the housing crisis and recession during this time period leads us to believe that zero emigration was not likely, especially for recently-arrived and Mexican-born immigrants.[3] Likewise, emigration rates for which P_{t2} is estimated using 2008 data are relatively higher than all other rates estimated for each subpopulation, which increases our concern about estimates based on 2008 data.

[3] Our results also include negative emigration rates for earlier arrivals and non-Mexicans based on 2007 and 2010 ACS data. Unlike the rates based on the 2008 ACS, zero falls within the pseudo-margin of error of the negative estimate in both cases. Low emigration is more plausible for these foreign-born subpopulations, which tend to be more settled and remain in the United States, as well.

A temporary change in data collection operations led to a relatively high number of missing values, and thus high imputation rates, in the 2008 ACS sample.[4] Given our results, we suspect the change may have increased non-sampling error in 2008 data and leads to downwardly biased estimates of the foreign-born population. Low population estimates, in turn, may be causing biased emigration rates. As we find in our results, if P_{t1} is underestimated, the residual calculation will result in a lower than expected, if not negative, emigration rate. An underestimate of P_{t2}, however, will result in higher than expected residual and emigration rate. Given the uncertainty in estimates based on 2008 data, we proceed by excluding such rates from the analyses and discussion that follows.

When we exclude emigration rates based on 2008 data, there is not a lot of variation in rates estimated for each subpopulation. Rates for the total foreign-born population range between 0.25 and 0.95, with only 0.25 (the 2007-2010 three-year rate) and 0.95 (the 2006-2009 three-year rate) differing by more than the pseudo-margins of error for the two rates.

In terms of differences across foreign-born subpopulations, our estimated emigration rates vary in expected directions. Recently-arrived immigrants emigrate at relatively higher rates (0.8 to 1.8) than earlier arrivals (0.0 to 0.4, assuming a negative rate implies zero). And Mexican-born immigrants leave the country at higher rates (1.24 to 2.16) than immigrants who are not from Mexico (0.0 to 0.6).

The 2006-2010 residual-based emigration rates for the total foreign-born population are relatively lower than estimates based on other time periods, data sources, and estimation methods (Figure 1). In particular, our rates are lower than previous decennial-to-decennial residual estimates, which is somewhat surprising given our expectation that a shorter observation horizon will lead to rates that are comparable to non-residual-based estimates. We suspect the low rates are due to a more diverse foreign-born population in 2006 relative to 1960, 1980, or 1990. The foreign-born population in the 2000s was comprised of more immigrants who tend to settle permanently upon arrival and not return home such as Asians and Africans (Gibson and Jung 2006; Grieco et al. 2012). This may not have been the case in the 1970s when Caribbean, Mexican, and Central American immigration grew rapidly, and the foreign-born population was more prone to return migration. This highlights a need to estimate emigration separately for foreign-born subpopulations that are known to have different migration and settlement patterns. Rates based on the residual method in general appear to be comparable to Schwabish's (2011) rates using administrative data but much lower than Van Hook et al.'s (2006).

[4] See note "2008 ACS Failed Edit Follow-up Operation" at http://www.census.gov/acs/www/data_documentation/user_notes/

Figure 1. Selected Rates of Foreign-Born Emigration, Total Foreign-Born Population in the United States, 1960 – 2010

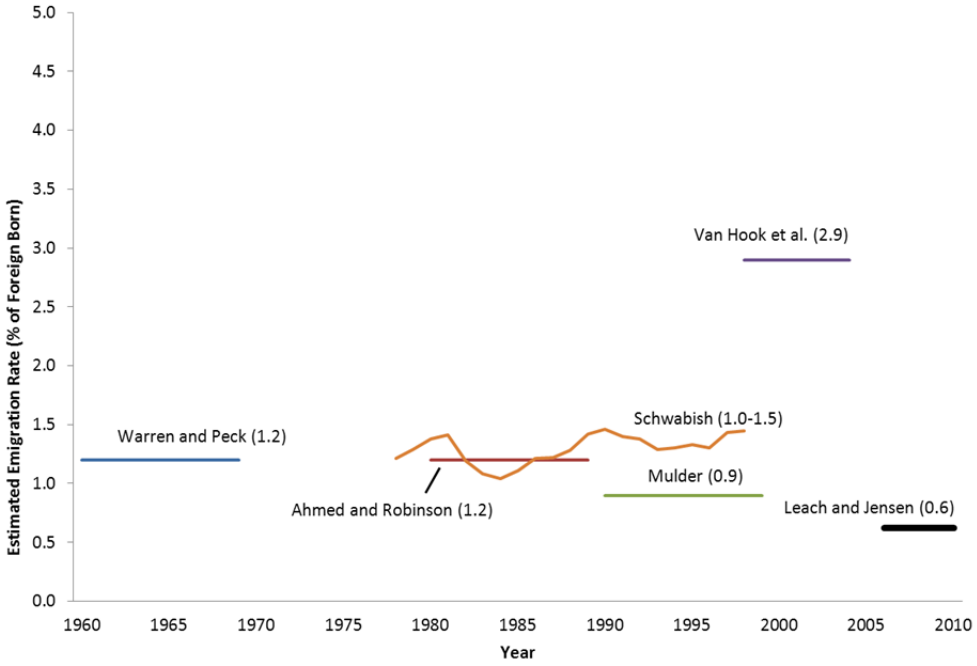
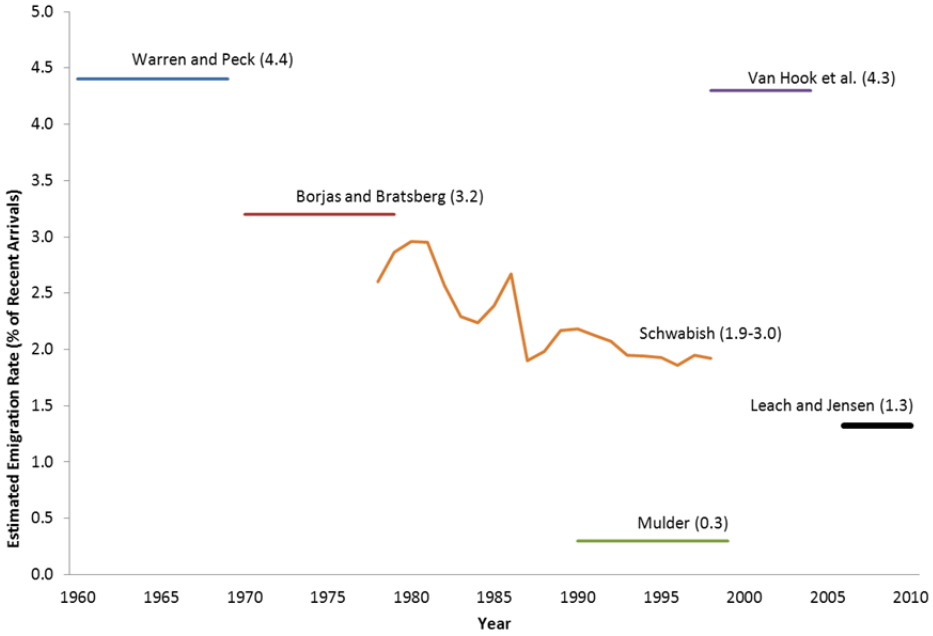


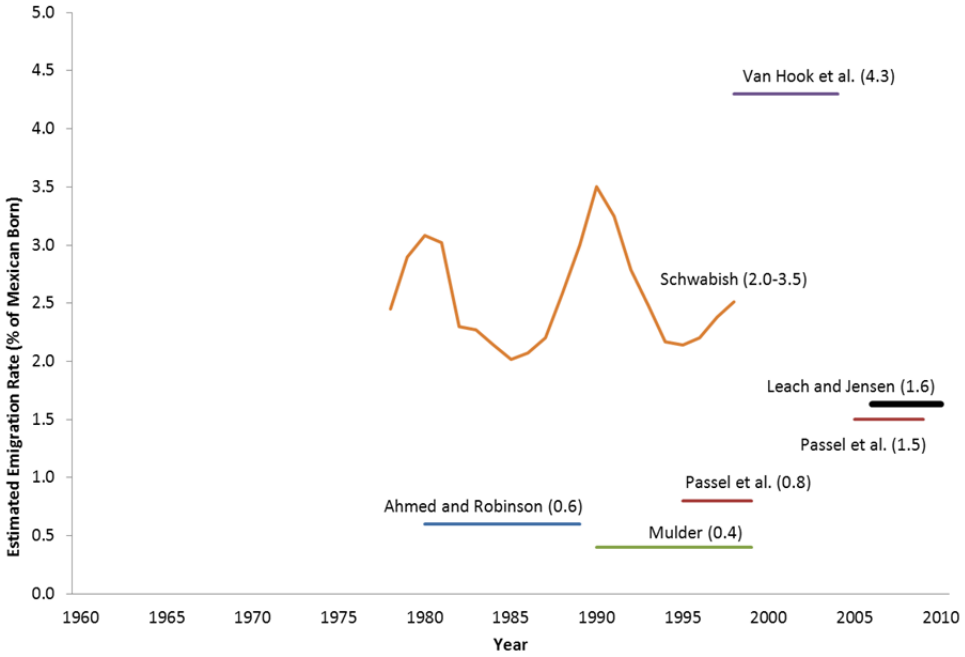
Figure 2 compares emigration rates for recently-arrived immigrants. The 2006-2010 estimated rates for recent arrivals are relatively lower than residual-based rates of both Warren and Peck (1980) and Borjas and Bratsberg (1996) but not those estimated by Mulder (2003). Again, it is difficult to ascertain whether differences are due more to methodological differences or change in the composition of the recently-arrived foreign-born population over time. Immigration in the 1950s and 1960s began to grow relative to historically small flows in the 1930s and 1940s (Gibson and Jung 2006). The growth was driven by immigration from Mexico and included many circular migrants (Bean and Stevens 2003; Massey, Durand and Malone 2002). More recently, immigration flows come from across the globe with some national origin groups settling permanently upon arrival while others emigrate with much greater propensity (Bean and Stevens 2003; Gibson and Jung 2006). Schwabish (2011), who calculates annual rates over 20 years, shows declining rates through the 1980s and 1990s.

Figure 2. Selected Rates of Foreign-Born Emigration, Foreign-Born in the United States for 10 Years or Less, 1960 – 2010



Unlike for the total foreign-born population, our 2006-2010 emigration rates for Mexican-born immigrants are relatively higher than previous residual estimates (Figure 3). It also is notable that our Mexican-born rates are more closely aligned with estimates based on other data and methods. While there is much variation in non-residual-based estimates, our Mexican-born emigration rates fall between rates based on Mexico Census data (Passel, Cohn and Gonzalez-Barrera 2012) and administrative data from the Social Security system (Schwabish 2011). These results support our expectation that residual-based emigration rates calculated using relatively shorter observation horizons produce better estimates, at least in terms of being more aligned with estimates based on other data and methods. This appears to be particularly true for a foreign-born subpopulation that has a relatively high propensity of return migration.

Figure 3. Selected Rates of Foreign-Born Emigration, Mexican-Born Population, 1960 – 2010



In spite of the promising results for Mexican-born emigration rates, comparisons between 2006-2010 residual estimates and previous decennial-to-decennial estimates may be confounded by historical context and population composition in addition to length of an observation horizon. The composition of the foreign-born population in the United States continues to change with respect to national origins that have different propensities to emigrate (Grieco et al. 2012; Jensen and Arenas-Germosen 2012). And historical events such as the terrorist attacks in 2001 and the 2008-2009 economic recession may also affect differences in propensity to emigrate across time periods. To investigate further the effect of length of observation horizon, we compare the 2006-2010 emigration rates to 2000-to-2010 emigration rates calculated using Census 2000 data and 2009 and 2010 ACS data. As noted above, the Census Bureau's current method of estimating foreign-born emigration relies on Census 2000 and ACS data. Also, we make comparisons both by recency of arrival (<=10 years, >10 years) and by place of birth (Mexico, non-Mexico) to simulate the Census Bureau's method of calculating emigration rates separately for the four foreign-born subpopulations.

Table 4 shows ACS-to-ACS and Census-to-ACS based emigration rates. In general, both methods generate emigration rates that differ between foreign-born subpopulations in expected ways. Earlier arrivals, whether Mexican or non-Mexican, tend to emigrate at lower rates relative to recent arrivals of similar national origin. It is also evident that the Mexican-born population emigrates at relatively higher rates than other foreign-born groups, just as we expected.

Table 4. Estimated Annual Rates of Foreign-Born Emigration Based on 2006 – 2010 ACS Five-Year Data and Census 2000 and 2009-2010 ACS Single-Year Data.

				Non-Mexican, > 10 Years in US		Non-Mexican, <= 10 Years in US		Mexican, > 10 Years in US		Mexican, <= 10 Years in US	
Data Sources	t1	t2	Years in Residual	Estimate (%)	Pseudo-MOE[1]	Estimate (%)	Pseudo-MOE[1]	Estimate (%)	Pseudo-MOE[1]	Estimate (%)	Pseudo-MOE[1]
ACS-to-ACS											
	2006	2009	3	0.3	0.34	1.2	0.59	0.7	0.59	**3.0**	0.93
	2007	2009	2	0.1	0.52	0.4	0.92	1.2	0.92	**3.6**	1.39
	2006	2010	4	0.1	0.24	0.4	0.48	0.2	0.49	**2.6**	0.75
	2007	2010	3	-0.2	0.33	-0.3	0.67	0.3	0.67	**3.0**	0.98
Census-to-ACS											
	2000	2009	9	0.2	0.05	1.4	0.09	0.8	0.09	**1.5**	0.12
	2000	2010	10	0.3	0.05	1.0	0.07	0.8	0.08	**1.4**	0.12

Source: Author's calculations using the 2006-2010 five-year ACS unpublished data for ACS-to-ACS rates & Census 2000 and 2009 and 2010 ACS single-year files for Census-to-ACS rates.

[1] It is difficult, if not impossible, to calculate a true margin of error for an emigration rate based on the residual method. We construct a "likely" range of values, which we refer to here as a pseudo-margin of error, based on 90 percent confidence intervals of the population estimates used in the residual calculation.

There are key differences between Census-to-ACS and ACS-to-ACS emigration rates, however. As discussed above, we expect a shorter observation horizon to be particularly important in estimating emigration for the recently-arrived Mexican-born population. ACS-to-ACS emigration rates for this population vary between 2.6 and 3.6 whereas Census-to-ACS rates range from 1.4 to 1.5. The difference appears to support our expectation that length of the residual period matters for residual-based emigration estimates and that a longer period biases downward annual emigration rates.

In contrast to rate differences for Mexican-born recent arrivals, the range of estimates produced by the ACS-to-ACS method for non-Mexican-born and Mexican-born earlier arrivals overlap the range of estimates from the Census-to-ACS method. Each of these groups emigrate at rates substantially lower than recently-arrived Mexican immigrants. The length of a residual period may not matter when return migration is less prevalent.

Conclusion

The primary objective of this paper is to assess the feasibility of using annual data from the American Community Survey to estimate foreign-born emigration from the United States. Typically based on subsequent decennial censuses, the most common critique of the residual method is that the measure excludes recently-arrived immigrants who are most prone to emigrate. When immigrants return home, they typically do so in the first few years after arriving in the United States (Massey, Durand and Malone 2002; Van Hook et al. 2006).

We use restricted micro data from the 2006-2010 ACS five-year file to produce rates of foreign-born emigration based on 2-, 3-, and 4-year residual periods. We first estimated rates for the entire foreign-born population, by time in the United States, and by Mexican origin. When the emigration rates based on 2008 data are excluded, ACS-to-ACS emigration estimates vary in expected directions. More recent arrivals emigrate at higher rates than earlier arrivals, which is consistent with the importance of time in country for processes of permanent settlement (Van Hook and Zhang 2011; Van Hook et al. 2006). ACS-to-ACS estimated rates also show that the Mexican-born population emigrates at much higher rates than non-Mexican-born immigrants, which also is well documented in the literature on international migration to and from the United States (Massey, Durand and Malone 2002; Massey and Singer 1995).

Comparing emigration rates within each subpopulation reveals potential susceptibility of ACS-to-ACS residuals to sampling and non-sampling error. In particular, emigration rates based on 2008 data appear to be biased due to

irregular data collection methods in that year. The direction of the bias depends on whether 2008 data are used to estimate the foreign-born population at the beginning or end of a residual period. This finding suggests the importance of calculating multiple rates based on different samples and varying residual periods when using the ACS to estimate emigration. Given the uncertainty surrounding the 2008-based rates, we proceeded with the analysis by excluding rates based on 2008 data.

Comparisons between ACS-to-ACS emigration estimates to previously published emigration rates reveal that ACS-to-ACS rates tend to be lower than previously-published estimates for the total foreign-born population (Figure 1) and foreign-born recent arrivals (Figure 2). ACS-based emigration rates for the total Mexican-born population, however, are relatively higher than previous residual-based estimates and fall within the range of estimates based on other data and methods (Figure 3). Given Mexican immigrants' known propensity to return home, a shorter residual period appears to address the main critique of the residual method and enhances the validity of a residual-based measure. When comparing rates based on different data, methods and time periods, however, differences also could be due to economic, social, and historical contexts and population composition.

To minimize the effects of context and composition, we compare ACS-to-ACS emigration rates to emigration rates based on Census 2000 and ACS data from 2009 and 2010. This provides a comparison for rates based on similar time periods but with different residual lengths. Emigration rates for the recently-arrived Mexican-born population were the only rates that differ significantly. This finding provides additional support for the notion that shortening a residual period enhances measurement of emigration, especially for a group with a relatively high propensity to emigrate.

To conclude, the present analysis shows that reducing the length of time use to calculate an emigration residual likely enhances estimates of foreign-born emigration.

References

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

Bean, F.D.and G. Stevens. 2003. *Americas Newcomers and the Dynamics of Diversity*. New York: Russell Sage Foundation.

Bhaskar, R., B. Arenas-Germosen, and C. Dick. 2013. "Demographic Analysis 2010: Sensitivity Analysis of the Foreign-Born Migration Component." Population Division Working Paper No. 98. Washington, DC: U.S. Census Bureau.

Bhaskar, R., S. Rastogi, and A. Kennedy-Puthoff. 2008. "Foreign-Born Emigration: Estimates and Rates." in *Annual Meetings of the Population Association of America*. New Orleans, LA.

Borjas, G.J.and B. Bratsberg. 1996. "Who leaves? The outmigration of the foreign-born." *Review of Economics and Statistics* 78:165-176.

Gibson, C.and K. Jung. 2006. "Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000." Population Division Working Paper No. 81. Washington, DC: U.S. Census Bureau.

Grieco, E., E.N. Trevelyan, L. Larsen, Y.D. Acosta, C. Gambino, G.P. De La Cruz, T. Gryn, and N.P. Walters. 2012. "The Size, Place of Birth, and Geographic Distribution of the Foreign-Born Population in the United States: 1960 to 2010." Population Division Working Paper No. 96. Washington, DC: U.S. Census Bureau.

Jasso, G.and M.R. Rosenzweig. 1982. "Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States." *Demography* 19:279-290.

Jensen, E.B.and B. Arenas-Germosen. 2012. "Recent Trends in the Racial and Ethnic Composition of Immigrant Flows." Presented at Population Association of America Annual Meeting, May 4, 2013, San Francisco, CA.

Massey, D.S., J. Durand, and N. Malone. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*. New York: Russell Sage Foundation.

Massey, D.S.and A. Singer. 1995. "New Estimates of Undocumented Mexcian Migration and the Probability of Apprehension." *Demography* 32:302-313.

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

Passel, J.S., D.V. Cohn, and A. Gonzalez-Barrera. 2012. "Net Migration from Mexico Falls to Zero—and Perhaps Less." *Pew Hispanic Center Report*. Available online at http://www.pewhispanic.org/2012/04/23/net-migration-from-mexico-falls-to-zero-and-perhaps-less/.

Rendall, M.S., P. Brownell, and S. Kups. 2011. "Declining Return Migration from the Unites States to Mexico in the late-2000s Recession: A Research Note." *Demography* 48(3):1049-1058.

Riosmena, F. 2004. "Return Versus Settlement Among Undocumented Mexican Migrants." Pp. 265-280 in *Crossing the Border: Research from the Mexican Migration Project*, edited by J. Durand and D.S. Massey. New York: Russell Sage Foundation.

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

Van Hook, J.and W. Zhang. 2011. "Who stays? Who goes? Selective emigration among the foreign-born." *Population research and policy review* 30(1):1-24.

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

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

Woodrow-Lafield, K.A. 1996. "Emigration from the USA: Multiplicity Survey Evidence." *Population research and policy review* 15:171-199.