How do the ACS five-year migration data compare to the 2000 Census migration data?

Megan Benetsky
William Koerber

Journey to Work and Migration Statistics Branch
Social, Economic, and Housing Statistics Division
Working Paper No. 2012-13
U.S. Census Bureau

Presented at the Annual Meeting of the Population Association of America,
San Francisco, CA May 3-5, 2012

This report is released to inform interested parties of ongoing research and to encourage discussion. The 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

From 1940 to 2000, the long form of the Decennial Census asked respondents about their migration patterns. The Census Bureau provided the public with an array of origin-destination migration flow data products, over the decades. The county-to-county migration flows, in particular, have continued to be a staple product. For Census 2000, a migration DVD with counts of movers was produced that contained a county-to-county flow table, and tables of mover counts by characteristics.

After Census 2000, the long form data was replaced by the American Community Survey (ACS). The 2010 Census and future decennial censuses have no migration data, making the ACS a primary source of migration data. ACS started collecting data in 1996 in four test sites. The scope of the survey grew and beginning in 2005 the survey sampled housing units in all counties in the U.S. and Puerto Rico. Group quarters (e.g., college dormitories, prisons, nursing homes, military barracks) were added to the sample in 2006.

One major difference between the Census and the ACS migration questions is the time reference for the question asked about a between the previous residence and the current residence. Census 2000 asked where the person lived 5 years ago (as of April 1, 1995) while the ACS asks where the person lived 1 year ago. The time period was changed to reflect the on-going data collection of the ACS, and allows for annual estimates of geographical mobility.

The Census Bureau released county-to-county migration estimates for the first time using the ACS in March 2012[1]. The working paper and data release provide the county-to-county flow and mover counts. This paper further analyzes the relationship of ACS migration data compared with the 2000 Census data. Because the ACS is the new source for migration data, it is important that these estimates are sound. This analysis looked for stability in the numbers, but also for reasonable changes in the data. As the estimates are determined reasonable, then mover characteristics are provided using ACS data.

We are expecting to find a larger number of movers in the 2000 Census data because it includes movers who moved between 1 and 5 years ago. We are also predicting that most of the county-to-county migration flow ratios will be relatively similar, though some outliers are expected.

[1] For more information on the 5-year ACS county-to-county migration estimates, see the working paper and the original tables here: http://www.census.gov/hhes/migration/data/acs/county-to-county.html.

DATA

The two surveys analyzed in this paper asked two different migration questions. The original question in the 2000 Census asks where the respondent lived 5 years ago. The question in the ACS now asks where the respondent lived 1 year ago. This presents some challenges to our data comparison.

The migration question from the 2000 Census spans 5 years, and therefore includes movers who moved 1, 2, 3, 4, and 5 years ago, while the 2005-2009 ACS question picks up people who only moved less than one year ago from the time they are surveyed. Due to this question change, the ACS is more likely to pick up temporary movers like college students than the 2000 Census. It is important to note that although the 2005-2009 ACS is a 5-year dataset, it is a 5-year period estimate using 1-year datasets. While it is not meant to approximate the 2000 Census data that asked about about moves in the last 5 years, it is the best data to use to check how "good" the ACS estimates are at capturing movement.

To illustrate the difference between the 1-year ACS estimates and the 5-year ACS estimates, Table 1 shows the number of people who moved into and left Los Angeles County, California in 2005, 2006, 2007, 2008, and 2009 are shown in the table below. The average number of movers entering Los Angeles County from 2005 to 2009 is 206,551. The ACS 5-year estimates indicate that there were 212,882 movers entering Los Angeles County during that time period. We know that the 5-year estimates are not five years of aggregated data, but rather a 5-year period estimate from 2005-2009.[2,3]

Table 1. ACS 1-Year Estimates 2005, 2006, 2007, 2008, 2009 vs ACS 5-Year Estimates 2005-2009, Los Angeles County, California

	Movers In	Movers Out
2005 ACS 1-Yr	191,159	365,203
2006 ACS 1-Yr	208,704	441,700
2007 ACS 1-Yr	210,432	391,845
2008 ACS 1-Yr	210,863	314,630
2009 ACS 1-Yr	211,595	323,968
Total	1,032,753	1,837,346
Average 2005-2009 1-Year Estimates	**206,551**	**367,469**
2005-2009 5-Year ACS Estimate	**212,882**	**372,331**

[2] The estimates across years may not be statistically different from one another.

3 For more information on how to interpret 5-year ACS estimates: www.census.gov/acs/www/Downloads/survey_methodology/Chapter_11_RevisedDec2010.pdf.

This analysis presents several figures and tables.[4] The first figure is a graph showing the distribution of coefficients of variation. The second figure is a scatterplot of the number of movers in the 2000 Census and the 2005-2009 ACS, using the smaller sample. The third figure shows the correlation and regression values for the entire sample of flows. The fourth figure shows the correlation and regression values for a sample including only the counties with at least 500 flows.

As the ACS county migration data were determined to be accurate estimates, counts were presented. This paper provides four tables of counts of county flow pairs and mover counts for both origin and destination counties. Additionally, tables with the following variable crosses and combination are discussed; age, sex, and race/Hispanic origin are shown independently, along with age by sex, age by race/Hispanic origin, sex by race/Hispanic origin, and age by sex by race/Hispanic origin.

COEFFICIENTS OF VARIATION

Coefficients of variation were calculated for each county-to-county flow pair. The percents were categorized as less than or equal to 10 percent variation, greater than 10 percent to 20 percent variation, greater than 20 percent to 30 percent variation, and greater than 30 percent variation and higher. Thirty percent variation is an important threshold determined by the Statistical Quality Standard F1-5 of the Census Bureau where "Serious data quality issues related to sampling error occur when the estimated coefficients of variation (CV) for the majority of the key estimates are larger than 30 percent."[5] The number in each county-to-county flow pair was also categorized and crossed by the CV categories in order to check that larger flows had little variation and smaller flows had more variation, as expected. Limiting the sample to flows that met these quality standards not only made for a better analysis, but also brought the sample size down from 238,435 to 6,295 flows. The analysis of these flows was reasonable even though they accounted for only about 2.6% of the total flows; they also accounted for about 40% of the total movers included in these flows.

In Figure 1, the categories with the smallest number of flows had the largest CVs; 91%-100% of the flows had CVs that were at least 30%. Categories with the largest number of flows had smaller CVs. However, in the category of 500-999 flows, the CV percent goes below 50%. This met the Statistical Quality Standard because the majority of the estimates had CV's less than 30%.

[Figure 1 about here]

[4] Movers from abroad are not included in this analysis.
[5] http://www.census.gov/quality/standards/standardf1.html

COMPARING MODELS

To compare the quality of the ACS data, correlations and regressions were estimated with the 2000 Census flows as the dependent variable and the 2005-2009 ACS flows as the independent variable. Pearson's r and the adjusted r-squared values were used to determine a good fit and consistency between the data. The results indicated that the pattern of flows in the ACS data are highly comparable to the patterns seen in the 2000 Census migration estimates.

The correlation for the full sample (807,099 flows[6,7]) showed that the flows in the 2005-2009 ACS were highly correlated with the 2000 Census, with a Pearson's r of about 0.94. When the 2005-2009 ACS flows were regressed on the 2000 Census flows, the ACS flows accounted for about 89.0% of the flows in the 2000 Census. Both of these statistics indicated that the ACS flow data were a good estimate of migration in relation to the 2000 Census data.[8]

The second model used only the counties associated with 500 or more migration flows. In this model, 4,297 flows have a Pearson's r of about 0.94. This smaller sample of the two surveys was correlated at almost the same magnitude as the full sample. This was not surprising given that despite the large decrease in sample size, much of the variation was removed when the sample was limited. The regression model showed that about 88% of the ACS flows predict the 2000 Census flows. This was slightly less than the adjusted r-squared value in the model with the full sample, but it was still a very good fit.[9]

COUNTS

Flows Out of Counties – The Origins

Tables 2 shows counts of "flows" for each origin and destination pair sorted from largest to smallest. A flow is a county migration pair, and a flow count shows how many county migration pairs with which that one county is associated. In Table 2, Maricopa County, Arizona had the largest number of "Flows Out" of the county. This meant that of the 3,141 counties in the U.S. that experienced outmigration, the movers leaving Maricopa County, Arizona moved to 1,156 other different destination counties. In other words, there were 1,156 "county flow pairs"

---

[6] Some ACS flows were collapsed to better reflect the 2000 Census vintage geography. For Alaska, Prince of Wales-Hyder Census Area was collapsed into Wales-Out Ketchikan Census Area, Wrangell City and Petersburg Census Area were combined, and Skagway Municipality and Hoonah-Angoon Census Area were combined. For Colorado, Broomfield County was collapsed into Boulder County. Also, several Virginia Independent cities and counties were combined. See http://www.census.gov/population/www/cen2000/ctytoctyflow/index.html for a complete list. This reduced the number of ACS flows from 238,435 to 236,162.

[7] The 807,099 is the number of unique flows from either ACS (236,162 total flows) or Census 2000 (735,531 total flows).

[8] See Appendix 1.

[9] See Appendix 2.

associated with Maricopa County, Arizona when it was the county where a move originated. Those 1,156 flows constituted about half of one percent of the total 238,435 county-to-county flows in the United States.

The remaining origin counties in the top ten after Maricopa County, Arizona included Los Angeles County, California; Cook County, Illinois; San Diego County, California; Harris County, Texas; Clark County, Nevada; Dallas County, Texas; Hillsborough County, Florida; Tarrant County, Texas; and Orange County, Florida. Aside from Cook County, Illinois, most of the origin counties are in the west and Florida. Some of the origin counties associated with the smallest number of flow pairs are Sioux County, Nebraska; Billings County, North Dakota; Kenedy County, Texas; Kalawao County, Hawaii; Grant County, Nebraska; Borden County, Texas; Loving County, Texas; Esmeralda County, Nevada; Storey County Nevada; and Yakutat City and Borough, Alaska. Neither Loup County nor McPherson County in Nebraska had any out flow migration.

Flows In to Counties – The Destinations

In Table 3, Maricopa County, Arizona had the largest "Number of Flows In." This meant that the movers who came to Maricopa County, came from 993 different counties in the United States. If there were a large number of flows in and out of a county, the movers entering and leaving the county were likely to be more dispersed throughout the country. If the number of flows in or out of a county was low, the counties the movers are coming from or going to are likely to be more concentrated. However, this is not always the case. In Figure 2, the map shows all of the origin counties for movers going to Maricopa County, Arizona. Figure 3 shows all of the origin counties for movers into Riverside County, California – a county that had a large amount of movers compared to Maricopa County, but had less than half of the county flows. The movers into Riverside County came from a more concentrated number of counties compared to movers into Maricopa County.

[Figures 2 and 3 about here]

The remaining destination counties among the top ten after Maricopa County, Arizona were Harris County, Texas; San Diego County, California; Los Angeles County, California; Cook County, Illinois; Bexar County, Texas; Clark County, Nevada; Tarrant County, Texas; Richland County, South Carolina; and Dallas County, Texas. Some of the largest flows for origin counties appeared as the largest flows for destination counties, but new counties like Bexar County, Texas and Richland County, South Carolina appeared as well. From these tables we learned that there are 3,142 sending counties and 3,141 receiving counties. Some of the destination counties with the smallest number of flow pairs were Petroleum County, Montana; Eureka County, Nevada; Blaine County, Texas; Loup County, Nevada; Oliver County, North Dakota; Liberty County, Montana; Baker County, Georgia; Kalawao County, Hawaii; Loving County, Texas; and Hinsdale County, Colorado. The last four were the smallest destination county with just one flow in each. Kenedy County, Texas had no in-migration.

Movers

While Tables 2 and 3 provided flows, Tables 4 and 5 show the actual number of movers from and to counties. Among the largest origin counties are Los Angeles County, California; Cook County, Illinois; Harris County, Texas; Maricopa County, Arizona; San Diego County, California; Dallas County, Texas; Orange County, California; Kings County, New York; New York County, New York; and San Bernardino County, California. By looking at movers rather than number of flows, we see that in addition to counties in California, Texas, Illinois, Nevada, and Florida, counties in New York also appeared. Looking at the bottom of the list, among the smallest number of movers are Terrell County, Texas; Arthur County, Nebraska; Hayes County, Nebraska; Borden County, Texas; Billings County, North Dakota; Blaine County, Nebraska; Grant County, Nebraska; Slope County, North Dakota; Loving County, Texas; and just four persons left Yakutat City and Borough, Alaska.

The number of movers entering each destination county are found in Table 5. Among the largest destination counties for movers are Los Angeles County, California; Maricopa County, Arizona; Harris County, Texas; Cook County, Illinois; San Diego County, California; Riverside County, California; Dallas County, Texas; San Bernardino County, California; Orange County, California; and King County, Washington. For movers, the biggest destination counties were in California, Arizona, Texas, Illinois, and Washington. At the bottom of the list, among the smallest number of movers entering were McPherson County, Nebraska; Keya Paha County, Nebraska; Hayes County, Nebraska; Baker County, Georgia; Kalawao County, Hawaii; Loving County, Texas; Loup County, Nebraska; and just two people moved into Hinsdale County, Colorado; Liberty County, Montana; and Oliver County, North Dakota.

Figure 4 is a scatterplot showing the number of movers in the 2000 Census by the number of movers in the 2005-2009 for county-to-county pairs with 500 or more movers. Here, the positive correlation between the surveys found previously is illustrated. The point to the far right was the number of movers from Los Angeles County, California to Orange County, California, where there were 146,044 movers in the 2000 Census and 41,612 movers in the 2005-2009 ACS. The data point to the left of the Los Angeles County-to-Orange County represented the number of movers from Los Angeles County, California to San Bernardino County, California, where there were 135,657 movers in the 2000 Census and 48,456 movers in the 2005-2009 ACS.

[Figure 4 about here]

Another way to use these data is to analyze the net gains or losses. By comparing Tables 4 and 5, one can calculate a county's net number of movers. For example, Los Angeles County, California lost 372,331 people to other counties (Table 4) and gained 212,882 people (Table 5), producing a net migration loss of 159,449 people for these years. Even though Los Angeles County, California sent more people than it received, it is still the largest destination county in the U.S.

It is important to note the distinction between flow counts and mover counts. As discussed earlier, flow counts are representative of how concentrated the counties sending or receiving movers were, while mover counts were strictly a summation of the number of movers in and out of counties. Typically, counties with a large number of movers are also associated with a large number of flows. However, it is possible that a county with a relatively large number of movers did not have a large number of flows. This is especially true for counties that are historically[10] known for high levels of migration; they have a large number of movers coming from fewer counties. For example, in Table 4, counties like Orange County, CA; Kings County, NY; New York County, New York; Riverside County, CA; and San Bernardino County, CA were all in the top ten counties with the largest number of movers leaving those counties. However, none of these counties were in the top 10 largest origin flow counts; only San Bernardino County, CA and Orange County, CA were in the top 20. These counties with historically high levels of migration, like New York County, NY, still had a large number of people entering (or leaving), but from (or to) a more concentrated number of counties.

CHARACTERISTICS

Unlike the flow and mover counts that had no restrictions on showing the complete estimates, the characteristics tables are held to a stricter level of disclosure avoidance.[11] The tables provided only include flows that had at least 3 movers from 3 different households, as per the Census Bureau's restriction guidelines. Characteristics were not analyzed for each county flow pair, but rather in the aggregate. The quality of the 5-year ACS estimates allowed for the analysis of these characteristics in accordance with disclosure avoidance policies. These policies will be applied to future releases of flow counts by characteristics.

The flow and mover counts across age categories are found in Table 6. The first and third columns are a count of all the county flow pairs and movers, respectively. According to the table there are 17,671 county flow pairs associated with 796,806 movers who are less than five years old. The first jump in the number of flows was for the ages 15-19 when flow counts go up to 27,554 and mover counts were 1,458,020. Flow and mover counts peaked at ages 20-24 with 34,054 flow pairs and 2,122,141 movers, and continued to decline until ages 85+ where there is a small increase in the number of flow pairs again.

Table 7 shows the flow and mover counts across sex. Men contributed to 48,570 flow pairs (6,594,184 movers) and women made up 45,999 flow pairs (6,041,115 movers). Taking age and sex together, most of the flow pairs were comprised of relatively young men, which is echoed throughout other migration research as well.

[10] http://www.brookings.edu/papers/2012/0320_population_frey.aspx
[11] For more information on disclosure avoidance, please see the working paper for the 2005-2009 5-year ACS county-to-county migration: http://www.census.gov/hhes/migration/data/acs/files/county-to-county/County_to_County_Mig_Working_Paper.pdf

Table 8 shows the flow and mover counts across race and Hispanic origin. All “Alone” categories describe races that are non-Hispanic. The number of flow counts for Hispanic movers was 18,616. In addition, there were 47,934 flow pairs for White Alone, 21,080 flow pairs for Black Alone, 3,752 flow pairs for American Indian/Alaskan Native, 9,572 for Asian Alone, 651 for Native Hawaiian/Pacific Islander Alone, 1,234 flow pairs for Other Alone, and 8,540 for Two or More Races. From these data, Whites Alone moved from a greater number of counties and in greater numbers than any other race or Hispanic origin, though this may be mostly due to the relative number of whites in the sample compared to the number of people of other races and Hispanic origin.

Table 9 crossed age and sex for flows and mover counts. In general, males moved in greater numbers and from a greater number of counties than females. This does not hold true between the ages of 15 and 19, where female movers outnumbered male movers by about 36,000. Males made up the largest number of movers between ages 20 and 59, until females again became more likely to move in the older ages, though the number of male and female movers between the ages of 55 to 59 are not statistically significant.

Table 10 crossed age and race/Hispanic origin. Here again, most of the movement for every race and Hispanic origin occurred between the ages of 15-29, though Asian Alone, Native Hawaiian and Pacific Islander Alone, and Other Alone also had a large percent of movers aged 30-34 (14.4 percent, 10.1 percent, and 11.9 percent, respectively). Hispanic or Latino and Two or More Races had a greater proportion of movement in the very young ages compared to the rest of the racial categories.

Table 11 crossed sex and race/Hispanic origin. For every race and for Hispanic or Latino, there were more male movers than female movers, though this is not statistically different for Other Alone, Native Hawaiian and Pacific Islander Alone, and Asian Alone. Male movers also came from a greater number of counties than female movers, except for Asian Alone. Female Asian Alone movers came from about 454 more counties than their male counterparts.

Table 12 crossed age, sex, and race/Hispanic origin. This table provided an even more detailed look at the general patterns observed from the previous tables. Some of the largest proportion of movers were between ages 15 and 29 across all race/Hispanic origin categories. However, males and females of Hispanic origin and Two or More Races had larger proportions of movers in the youngest ages compared to other race categories, as 9.0 percent of male movers of Hispanic origin were less than five years old versus 10.9 percent of Hispanic origin females.[12] In addition, for Two or More Races, 14 percent of males and 13.5 percent of females were less than five years old.

Another interesting pattern arose at the oldest age category. Across each race/Hispanic origin category, a larger proportion of females 85 and older moved compared to their male peers, except for Other Race Alone. Native Hawaiians and Pacific Islanders Alone were the largest proportion followed by White Alone. However, the Native Hawaiian and Pacific Islanders group

[12] Though male Hispanics aged 0-5 are not significantly different from Other Alone race ages 0-5.

was not significantly different from White Alone, Black Alone, American Indian and Alaskan Native Alone, or Hispanic. White Alone was not significantly different from Native Hawaiian and Pacific Islanders Alone.

SUMMARY

This analysis has demonstrated that the ACS 5-year county-to-county migration estimates are comparable to the data obtained in the 2000 Census, the most historical source of small scale geographic migration data. Despite comparing two different surveys utilizing two different migration questions, there is congruence in the relative magnitude of county-to-county movers found between the surveys. This finding is essential to the future of migration research, as the ACS will be the survey of record for migration.

As the ACS 5-year estimates were well predicted by data from the 2000 Census, we provided estimates by age, sex, and race/Hispanic origin. The general pattern found in the data was that movers tend to be young and male. Movers of Hispanic origin or were Two or More Races had an even larger proportion of movers at very young ages. Females were more likely to move in the oldest ages compared to males across all race/Hispanic origin categories except Other.

Future releases of characteristics are planned using the 5-year ACS estimates. Starting with the release of the 2006-2010 ACS county/mcd-to-county/mcd migration flows, files will be released with selected characteristics: age, sex, and race/Hispanic origin. Additional characteristic cross-tabulations will be released in sequential years.

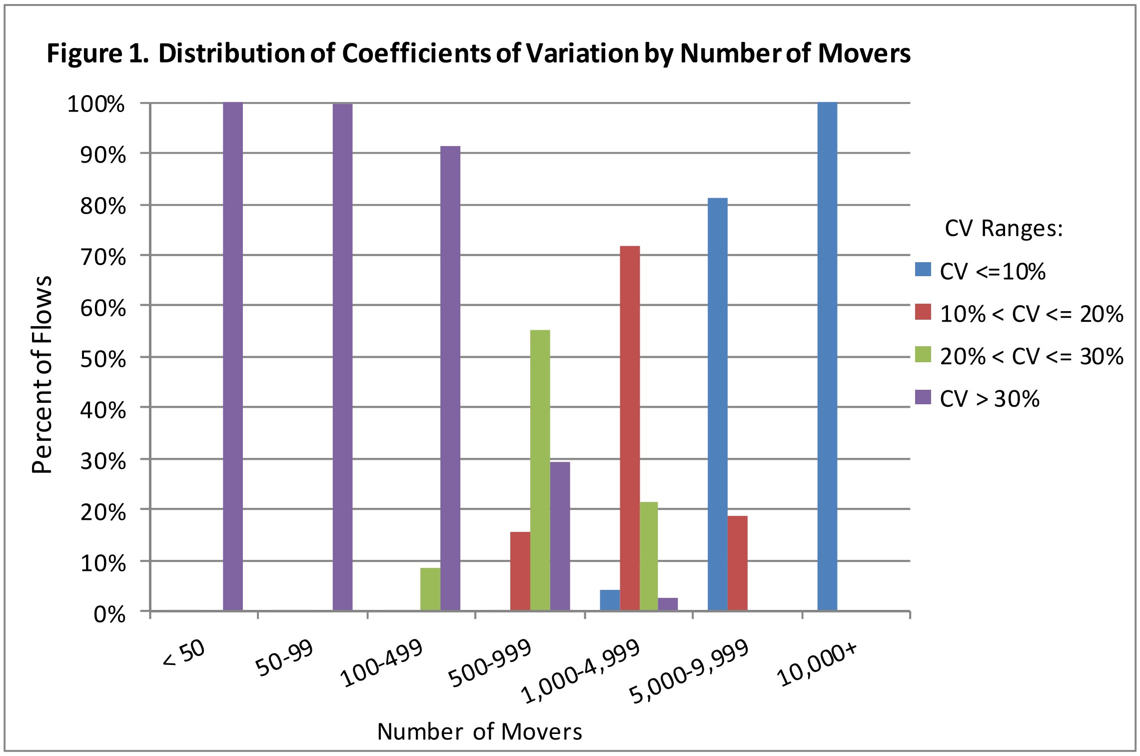
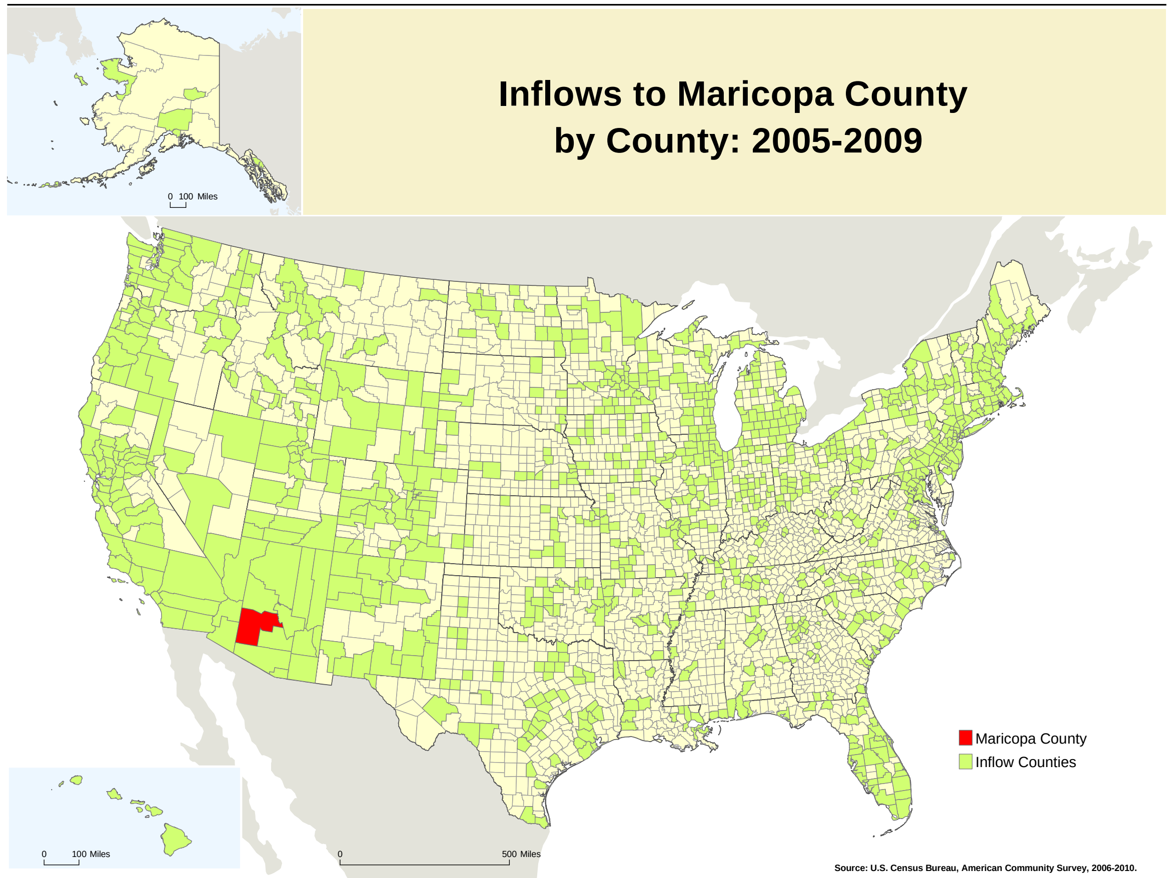


Figure 1. Distribution of Coefficients of Variation by Number of Movers
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percent of Flows
< 50
50-99
100-499
500-999
1,000-4,999
5,000-9,999
10,000+
Number of Movers
CV Ranges:
CV <=10%
10% < CV <= 20%
20% < CV <= 30%
CV > 30%

Figure 2.



Inflows to Maricopa County by County: 2005-2009
0 100 Miles
0 100 Miles
0 500 Miles
Maricopa County
Inflow Counties
Source: U.S. Census Bureau, American Community Survey, 2006-2010.

Figure 3.

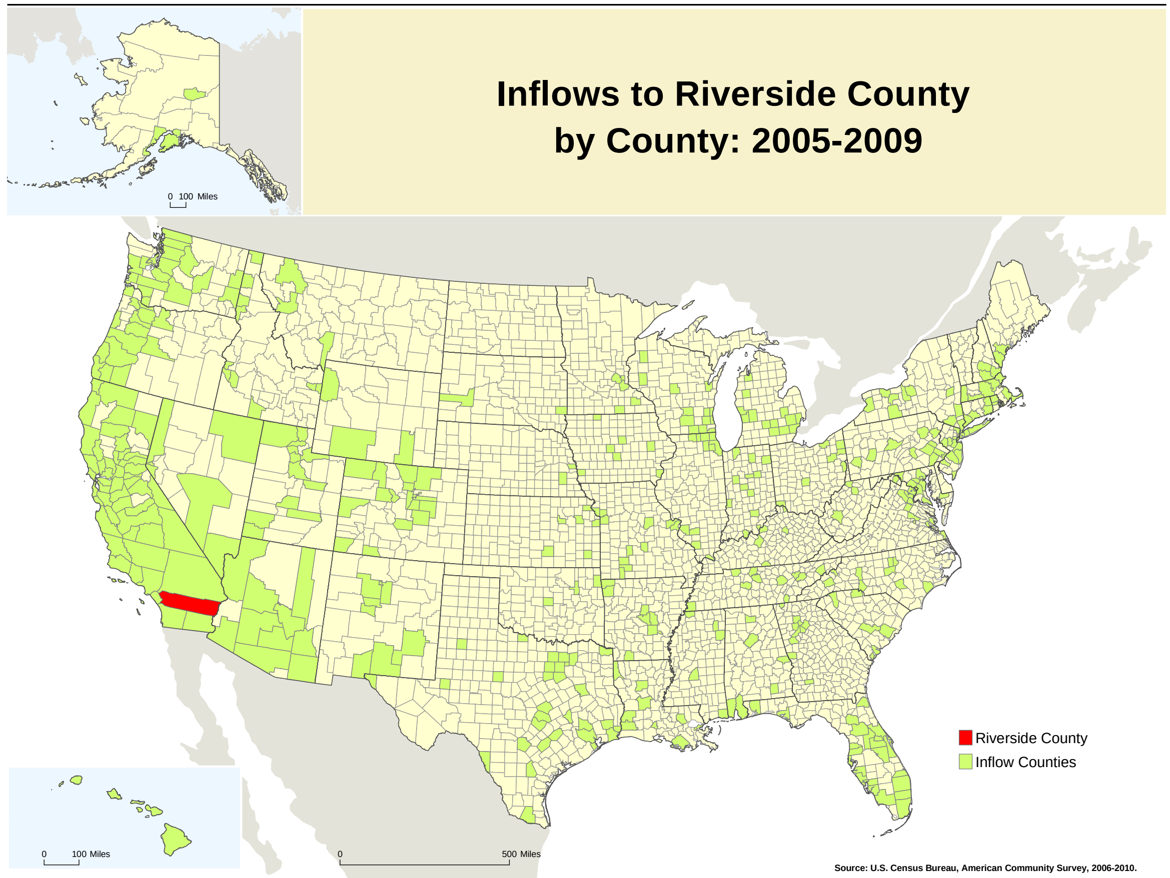
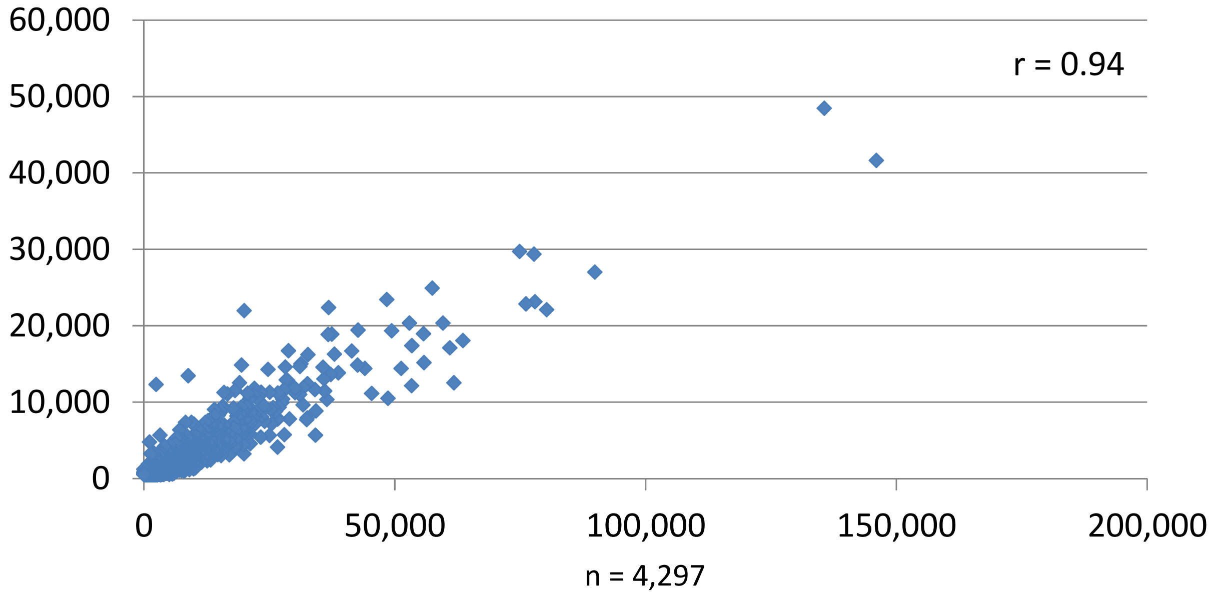

Inflows to Riverside County
by County: 2005-2009
0 100 Miles
0 100 Miles
0 500 Miles
Riverside County
Inflow Counties
Source: U.S. Census Bureau, American Community Survey, 2006-2010.

Figure 4. Number of County-to-County Movers in 2000 Census by 2005-2009 ACS, for Flows Greater than 500



Appendix 1.

Pearson Correlation Coefficients, Full Sample
N = 807,099

	Flow2000	Flow0509
Flow2000	1	0.94316 ***
Flow0509	0.94316 ***	1

**** Significant at the .001 level.*

Linear Regression of ACS 2005-2009 Flows, Full Sample

Intercept	1.30 ***
Flow2000	0.35 ***
N	807,099
Adjusted R^2	0.89

**** Significant at the .001 level.*

Appendix 2

Pearson Correlation Coefficients, 500+ Flows Sample
N = 4,297

	Flow2000	Flow0509
Flow2000	1	0.93856 ***
Flow0509	0.93856 ***	1

**** Significant at the .001 level.*

Linear Regression of ACS 2005-2009 Flows, 500+ Flows Sample

Intercept	296.24	***
Flow2000	0.34	***
N	4,297	
Adjusted R^2	0.88	

**** Significant at the .001 level.*