

Federal Reserve Bank of Minneapolis
Research Department

Interstate Migration Has Fallen Less Than You Think: Consequences of Hot Deck Imputation in the Current Population Survey*

Greg Kaplan and Sam Schulhofer-Wohl

Working Paper 681

November 2010

ABSTRACT

We show that the significant drop in the annual interstate migration rate between the 2005 and 2006 Current Population Surveys is a statistical artifact. The Census Bureau's imputation procedure for dealing with missing data before the 2006 survey year inflated the estimated interstate migration rate. An undocumented change in the procedure corrected the problem for the 2006 and later surveys, thus reducing the estimated migration rate. The change in imputation procedures explains 90 percent of the reported decrease in interstate migration between 2005 and 2006, and 42 percent of the decrease between 2000 (the recent high-water mark) and 2010. After we remove the effect of the change in procedures, we find that the annual interstate migration rate follows a smooth downward trend from 1996 to 2010. The 2007–2009 recession is not associated with any additional decrease in interstate migration relative to trend.

Keywords: Interstate migration; Mobility; Current Population Survey; Hot deck imputation; Missing data; Item nonresponse

JEL: J11, R23, C81, C83

*Kaplan: University of Pennsylvania. Schulhofer-Wohl: Federal Reserve Bank of Minneapolis and National Bureau of Economic Research. We thank Ellen McGrattan, Andrew Gelman and an anonymous commenter for helpful suggestions and Joan Gieseke for editorial assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

The low rate of migration between U.S. states has attracted much recent attention from demographers, policymakers, and the news media. Observers have often emphasized that, in addition to the long-run downward trend in interstate migration over the past half century, the migration rate fell sharply in 2006 from a relatively high plateau between 1999 and 2005. Figure 1(a) shows the annual interstate migration rate for the past decade, as calculated by the U.S. Census Bureau and published on its web page.¹ Taking note of these data, a front-page article in the *Washington Post* laments that “labor mobility has nearly ground to a halt in the past two years” and reports that “the recent sharp downturn” in interstate migration is making economists, including Assistant Treasury Secretary Alan B. Krueger, worry that low mobility will harm the nation’s recovery from the recession (Fletcher, 2010). Frey (2009), analyzing the Census Bureau data, calls the decline in interstate migration “dramatic” and notes that “the 1.6 percent interstate migration rate for the past two years was half the value exhibited in 1999–2000” (p. 4). And a recent International Monetary Fund report on the U.S. economy argues that “slower inter-state migration, likely related to the housing crash,” is raising unemployment (Batini et al., 2010, pp. 5–6).

Implicit in these discussions is the hypothesis that a decline in interstate migration reduces the economy’s ability to respond to adverse shocks. Whether that hypothesis is correct, and what government policies might constitute an appropriate response, are important research questions. However, before such questions can be addressed, a crucial first step is to obtain accurate data on what is actually happening to migration rates.

In this paper, we show that interstate mobility *has not* fallen dramatically in recent years. The migration rate has merely followed its long-term downward trend. Analyses that have found a sharp drop relative to trend have been based on the Census Bureau’s published rate, which the bureau calculates from the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS). In 2006, the Census Bureau made a seemingly minor change in its procedures for imputing missing data in the ASEC. This change in imputation procedures — *not* any actual change in migration patterns — is responsible for

¹“Geographical Mobility/Migration,” U.S. Census Bureau, accessed Oct. 28, 2010, <http://www.census.gov/population/www/socdemo/migrate.html>.

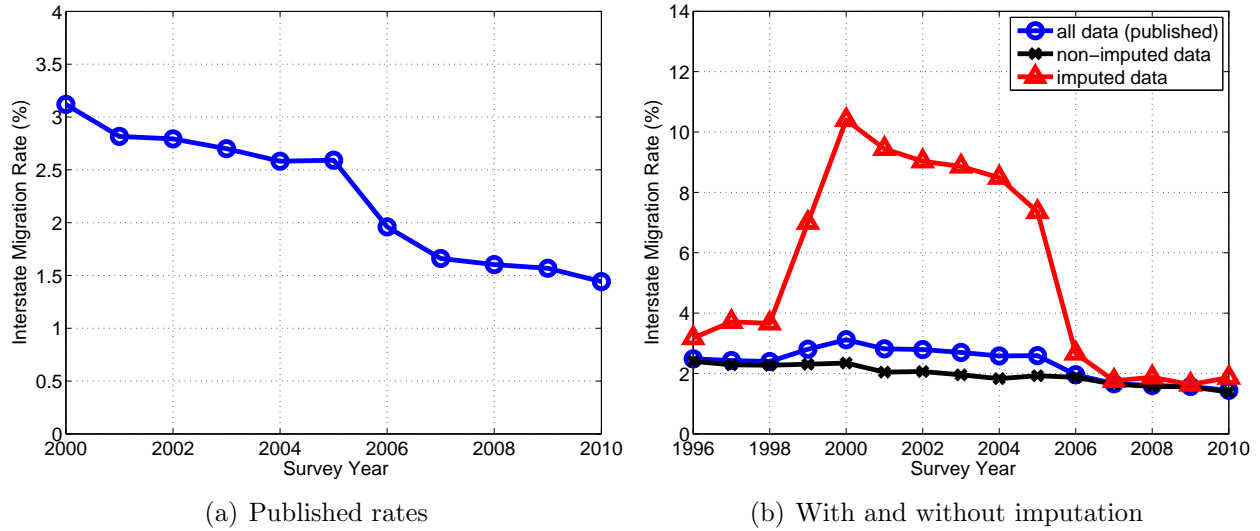


Figure 1: Rates of migration between states.

(a) Rate is weighted percentage of respondents at least one year old who lived in a different state one year ago, as calculated by the Census Bureau using all data in the Annual Social and Economic Supplement to the Current Population Survey. Source: 2000–2009 survey years from CPS ASEC tables at <http://www.census.gov/population/www/socdemo/migrate.html>; 2010 survey year from authors’ calculations from CPS ASEC. (b) Rate is weighted percentage of respondents at least one year old who lived in a different state one year ago. Source: Authors’ calculations from CPS ASEC.

much of the decline in reported ASEC migration rates. The change explains 90 percent of the reported decrease in interstate migration between 2005 and 2006, and 42 percent of the decrease between 2000 (the recent high-water mark) and 2010.

Figure 1(b) illustrates the problem. The figure shows the interstate migration rate for all ASEC respondents, and for those with original and imputed data separately. From 1996 to 1998 and from 2006 to 2010, the rate for respondents with imputed data is only slightly higher than the rate for respondents with original data, and the rate for all respondents is likewise very close to the rate for respondents with original data. But from 1999 to 2005, the interstate migration rate for respondents with imputed data is *three to five times* the rate for respondents with nonimputed data. Including the imputed data drives up the rate for all respondents, which is the rate that the Census Bureau publishes and that recent research, policy, and media reports have discussed.²

²After we noticed the discrepancy between imputed and nonimputed data, we corresponded with Census Bureau staff. They told us that the imputation procedure was changed in 2006 in such a way as to reduce

The paper proceeds as follows. Section 2 explains how the imputation procedures changed. Section 3 documents the effect of the change on migration rates estimated from the ASEC. Section 4 argues that the nonimputed data provide the most accurate guide to recent trends in interstate migration and uses independent datasets to check the robustness of our results. Section 5 discusses how our findings relate to the policy discussion on migration in the current recession and suggests some avenues for further research.

2. Imputation procedures for migration data in the ASEC

The ASEC, informally known as the March CPS because most of the data are collected in March, has been used to construct published annual migration data since 1948. It is thus the longest-running migration data series available for the United States. Unlike decennial census data, ASEC data can reveal short-run changes in migration patterns. Unlike the American Community Survey (ACS), which collects data year round, the ASEC collects data at a relatively precise point in time.³ Unlike data on tax return mobility from the Internal Revenue Service (IRS), the ASEC mobility data are part of a dataset with many covariates, allowing researchers to study the individual and household characteristics that influence mobility. And unlike panel data, ASEC data are not confounded by the difficulty of finding respondents when they move.⁴ Hence, the ASEC is a unique and invaluable resource for research on internal migration in the United States.

But as valuable as it is, the ASEC — like all surveys — must be analyzed with caution. In this case, the caution relates to missing data. The ASEC suffers from a significant amount of item nonresponse: cases where respondents answer some questions but not others. The Census Bureau can sometimes infer the correct answer from the answers of other household members. But often there is no information about the correct answer. In recent years, migration information was missing and could not be inferred from household members' answers

the interstate migration rate. This change is mentioned in a footnote in an unpublished Census Bureau working paper that compares the ASEC to the American Community Survey (Koerber, 2007, p. 14), but is not discussed on the web page where the Census Bureau tabulates migration rates, nor in any official ASEC documentation of which we are aware.

³The 2009 ASEC tells us the fraction of Americans who moved between February–April 2008 and February–April 2009, while the 2009 ACS provides only an average of one-year migration rates for intervals from January 2008–January 2009 to December 2008–December 2009.

⁴The ASEC measures mobility with retrospective questions: “Did this person live in this house or apartment 1 year ago?” and “Where did (reference person’s name/you) live one year ago?”

for 10 to 12 percent of ASEC respondents. It would be inappropriate to calculate migration rates using only the data from people who answer the questions, because those who answer may differ in many ways from those who do not answer. In principle, one could calculate bounds on the migration rate by assuming alternatively that all nonrespondents are migrants or that none are.⁵ However, because nonresponse is so common, the bounds would be so wide as to be almost useless. Some method of guessing or imputing nonrespondents' answers is therefore crucial for obtaining accurate estimates of migration rates.

The Census Bureau fills in missing values in the ASEC using a method known as hot deck allocation. The technical documentation for the Current Population Survey (U.S. Census Bureau, 2006, chap. 9) describes the method in detail. Briefly, respondents are assigned to cells based on observable characteristics. Which characteristics are used depends on what variable is being imputed. According to Koerber (2007), for imputing mobility variables the 2004 ASEC used cells based on census division of current residence, race, housing tenure, age, and armed forces status. Missing values are filled in one respondent at a time. If a respondent did not answer a question, the answer is filled in by copying the answer of the most recently processed respondent who falls in the same cell and who did answer the question. The person whose missing answer is filled in is known as a recipient; the person whose answer is used is known as a donor.

For migration data, the imputation procedure substitutes donors' answers for missing answers to two questions: whether the respondent lived in the same home one year ago and, if not, where the respondent lived one year ago. Once these variables are filled, additional variables are calculated that categorize movers as having moved within a county, between counties in the same state, between states, or from abroad, based on the distance between the respondent's current location and her (possibly imputed) location one year ago.

Using the most recently processed respondent as the donor to impute missing answers means that the order of processing can affect the results. Since 2006, respondents have been processed in geographic order. This ordering means that the donor usually lives near the recipient. Since long-distance migration is rare, the donor's location one year ago is also usually close to the recipient's current location. Thus, if the procedure imputes that the

⁵See Brown (1984) for an application of bounds to missing earnings data in the CPS.

recipient moved, it usually imputes a local move. Before 2006, the order of processing was geographic but within particular samples.⁶ (The ASEC consists of several samples, not one; see U.S. Census Bureau, 2006, chapter 11.) Therefore, on average, donors lived farther from recipients; donors' locations one year ago were also on average farther from recipients' current locations; and recipients were more likely to have imputed interstate moves.

In sum, the switch to geographic sorting in 2006 reduced the reported interstate migration rate by imputing fewer interstate moves and more local moves. For example, suppose a person in Philadelphia fails to answer the migration questions and is matched with a donor who moved, so that the nonrespondent is coded as a mover. If imputations are done in geographic order, the donor will probably also come from Philadelphia, and the donor's location one year ago was also probably near Philadelphia. The geographic procedure will thus usually impute that the nonrespondent made a local move. However, if imputations are not done in geographic order, the donor may come from farther away — Minneapolis, say — and the donor's location one year ago was probably near Minneapolis, not Philadelphia. The pre-2006 nongeographic procedure will thus impute that the nonrespondent made an interstate move, not because the nonrespondent is similar to a respondent who moved a long distance, but rather because the nonrespondent is similar to a respondent who moved a short distance in Minnesota.

We have not been able to determine why the imputed and nonimputed data began diverging in 1999 and diverged more in 2000. Samples were added to the CPS in 2001, but this would have led the series to diverge in 2001, not in 1999 and 2000. Census Bureau staff told us that the only change to the imputation procedure in 1999 was a change in the computer language in which the programs are written.

3. The effect of imputed data on estimated migration rates

We analyze migration rates in the 1996 to 2010 ASEC survey years. Data from the 1995 survey are not comparable because migration questions were asked in a different way, and in earlier years, the public-use data files either do not contain flags for imputed migration data

⁶E-mail to Schulhofer-Wohl from David K. Ihrke, Journey-to-Work and Migration Statistics Branch, Housing and Household Economic Statistics Division, U.S. Census Bureau, Oct. 21, 2010. See also Koerber (2007).

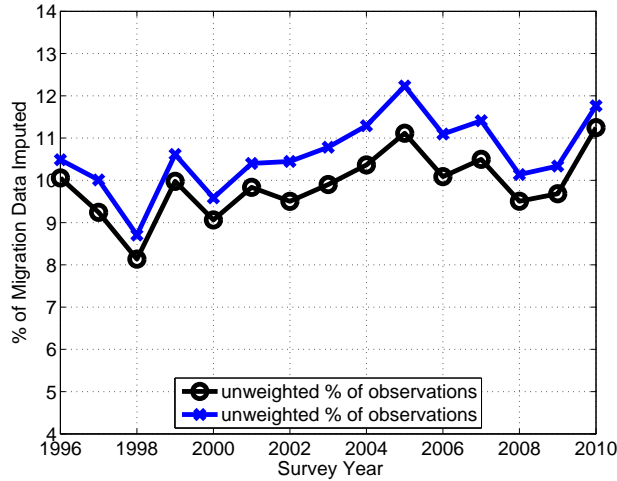


Figure 2: Migration imputation rates.

Rate is percentage of respondents at least one year old who had imputed migration data. Source: Authors’ calculations from CPS ASEC.

or show that virtually no migration observations were imputed. We do not report standard errors or confidence intervals because the ASEC sample is so large — more than 200,000 individuals in recent years — that the standard errors for aggregate statistics are minuscule.⁷

We consider a respondent to have imputed data only if the data came from hot deck allocation. We categorize respondents as nonimputed if their migration data were inferred from household members’ survey responses.⁸ Figure 2 shows the percentage of respondents

⁷In the 2010 ASEC, the standard error of the interstate migration rate, calculated using the replicate weights that accompany the public-use data file, is 0.05 percentage point for the full sample and for observations with original migration data, and 0.18 percentage point for observations with imputed migration data.

⁸Specifically, we categorize a respondent as having imputed migration data if migration status (whether the respondent lived in the same home one year ago) was hot deck allocated; if migration status was not allocated but the person is coded as a migrant and state of residence one year ago was hot deck allocated; if migration status was not allocated, the person is coded as a within-state or within-county migrant, and the county of residence one year ago was hot deck allocated; or if the person’s migration data were inferred from a householder, parent or spouse whose migration data in turn were hot deck allocated or inferred from yet another respondent whose data were hot deck allocated. For 1.6 percent of weighted observations, the imputation flags show that the migration data were inferred from another respondent, but this other respondent does not exist. This problem mainly occurs when an unmarried person has an imputation flag indicating “assigned from spouse.” We treat these unlinkable observations as having nonimputed data since there is no information on the data file to indicate that their data came from hot deck allocation. However, the migration rates in the unlinkable observations follow the same pattern as the rates in imputed data, so some researchers may prefer to treat the unlinkable observations as imputed. Classifying the unlinkable observations as hot deck allocated does not significantly change our findings; these results are available on request.

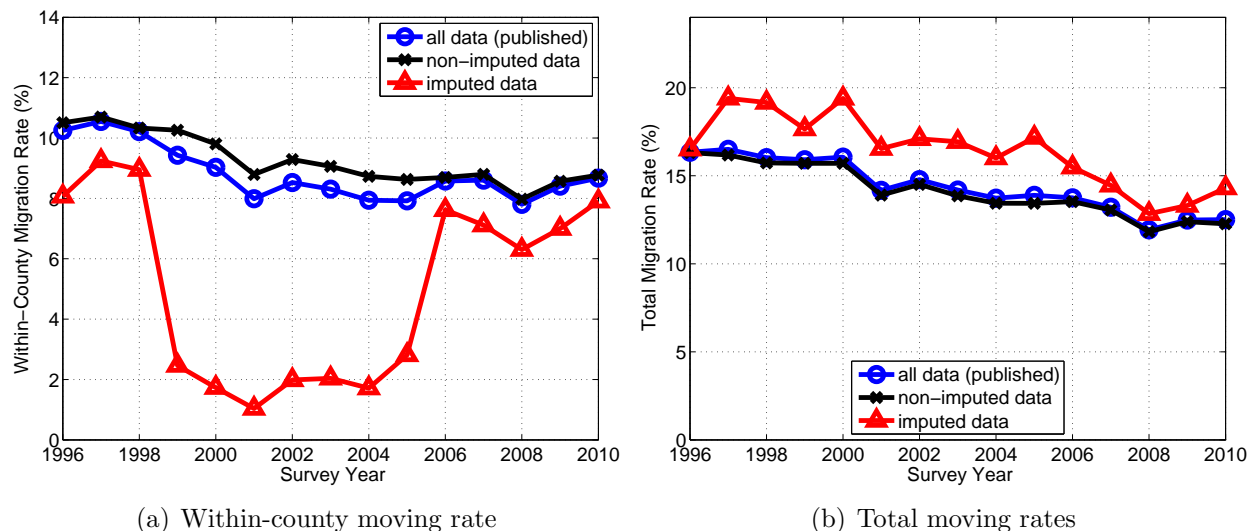


Figure 3: Moving rates.

(a) Rate is weighted percentage of respondents at least one year old who lived in a different home in the same county one year ago. (b) Rate is weighted percentage of respondents at least one year old who lived in a different home one year ago. Source: Authors' calculations from CPS ASEC.

who have imputed migration data in each survey year. The rate of imputed data is stable but quite high relative to the migration rate, illustrating the need to account for nonrespondents in some way.

In figure 1(b), we demonstrated that imputed data show a much higher interstate migration rate than nonimputed data from 1999 to 2005. We now examine how imputations affect the rate of moving within counties and the total migration rate for moves of all distances.⁹ Figure 3(a) shows that the change in imputation procedures had the opposite effect on within-county migration data compared with the effect on interstate migration data: From 1999 to 2005, the within-county migration rate was depressed among respondents with imputed data. Thus, the pre-2006 imputation procedure spuriously imputed long-distance moves that should have been local moves. Figure 3(b) shows that the change in imputation procedures had little effect on the total migration rate, because the decrease in imputed interstate moves in 2006 cancels out an increase in imputed within-county moves.

The comparisons of migration rates in imputed and nonimputed data do not, of course,

⁹The change in imputation procedures had little effect on rates of migration from abroad and between counties in the same state. These results are available on request.

prove that the pre-2006 imputation procedure led to an artificially inflated interstate migration rate and a reduced within-county rate. After all, the entire reason for imputing data is that respondents with missing data may differ systematically from those with nonmissing data. Ideally, we would show the effect of the change in imputation procedures by redoing the pre-2006 imputations using the sort order in effect since 2006. However, this calculation would require access to geographic and other data that the Census Bureau does not make public.

We nonetheless think there is a very strong case that the change in imputation procedures explains why the interstate migration rate was higher before 2006. First, this is the explanation Census Bureau staff gave when we brought the issue to their attention. Second, mechanically, the geographic sort in place since 2006 must lead to at least some reduction in imputed interstate moves. Finally, Koerber (2007) indicates that the pre-2006 sort order raised the interstate migration rate relative to a geographic sort used in the ACS.

4. What is the true trend in interstate migration?

We agree with the Census Bureau that the geographic sort that is currently in use is likely to produce the most reliable imputations of migration rates. Any other sort order will impute too many long-distance moves simply because some recipients will live far from their donors.¹⁰ The pre-2006 data likely overstate the rate of interstate migration and understate the rate of local migration.

The change in imputation procedures in 2006 means that comparisons of pre-2006 and post-2006 data do not accurately measure trends in interstate migration. However, because there may have been good reasons for using different sort orders to obtain better imputations of other variables in past surveys, we do not think it would necessarily be desirable for the Census Bureau to recalculate the pre-2006 imputations using the current procedure.¹¹ Fortunately, at least for analyzing the trend in interstate migration rates, there is a simple way to obtain accurate estimates for past years. Since 2006, the interstate migration rate including

¹⁰Alternatively, to reduce the sort order's impact on the results, one could impute the type of move (for example, interstate or within-county) instead of inferring the type of move from an imputed location one year ago.

¹¹The same sort order must be used for imputing all variables because some imputations are functions of other imputations.

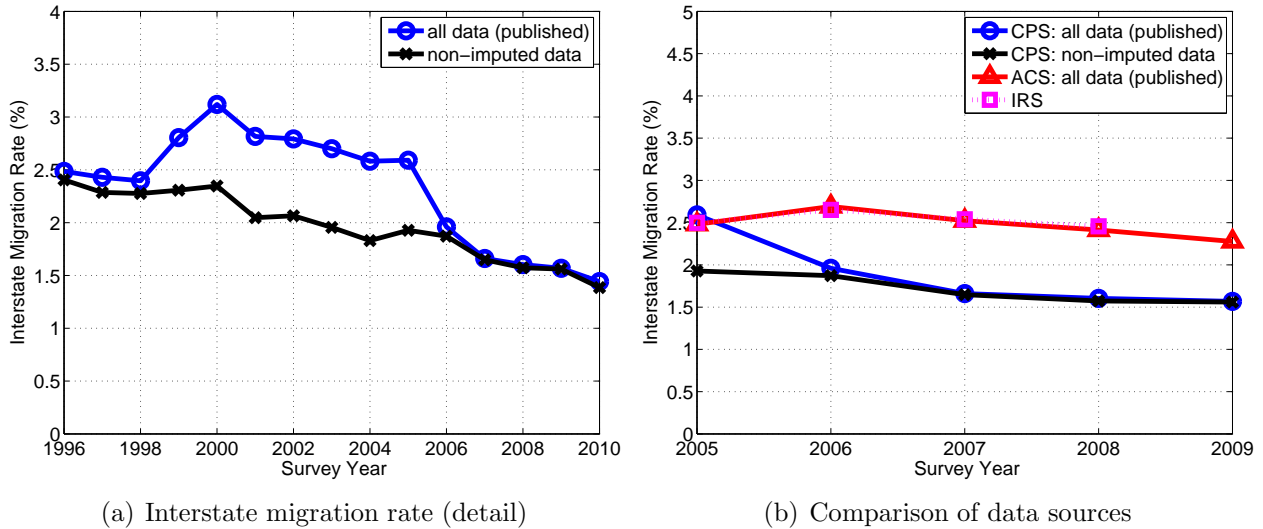


Figure 4: Rates of migration between states.

(a) Rate is weighted percentage of respondents at least one year old who lived in a different state one year ago. (b) CPS and ACS rates are weighted percentages of respondents at least one year old who lived in a different state one year ago. IRS migration rate is the percentage of tax returns that were filed in a different state the previous year, weighted by the number of exemptions claimed on the return, with previous year's location defined by the filing location of the primary taxpayer on the return, and with new tax filers excluded from the calculation. Source: Authors' calculations from CPS ASEC and IRS Statistics of Income data; ACS rates from Census Bureau's American FactFinder.

imputed data has been virtually identical to the rate using only nonimputed data. Figure 4(a) illustrates this point by reproducing figure 1(b) without the imputed-data migration rate, to show only the contrast between overall and nonimputed rates. Because the interstate migration rates using nonimputed data and using all data have been virtually identical for the past five years, we think that the rate using nonimputed data is a reliable guide before 2006 as well. Analysts interested in trends in the overall interstate migration rate should focus on the rate in nonimputed data, and doing so will remove the fluctuations induced by changes in the imputation procedure.

Figure 4(a) shows that, once we remove the effect of changes in the imputation procedure, the interstate migration rate has hewed closely to a smooth downward trend for the past 15 years. With imputations included, the rate peaked at 3.12 percent in the 2000 survey, fell to 2.59 percent by the 2005 survey, plummeted to 1.96 percent in the 2006 survey, and is now down to 1.44 percent. Without imputations, the rate was 2.35 percent in the 2000

survey, 1.93 percent in the 2005 survey, 1.87 percent in the 2006 survey, and 1.38 percent in the 2010 survey. Thus, the change in imputation procedures explains nine-tenths of the 0.63 percentage point drop from 2005 to 2006 and four-tenths of the 1.68 percentage point drop from 2000 to 2010.

We can also check how much the interstate migration rate has fallen in recent years by comparing the ASEC results with data from the ACS and the IRS. Neither the ACS nor the IRS data are exactly comparable to the ASEC data. The ACS tends to smooth out fluctuations in the migration rate because ACS data are collected year round rather than mainly in March. In addition, Koerber (2007) notes that the ACS tends to find more migrants than the ASEC because of differences in survey procedures.¹² Meanwhile, the IRS data cover only people who file income-tax returns, which means low-income people are underrepresented; track tax return mailing addresses rather than residence addresses; and do not measure migration at a precise date because tax returns enter the calculations if they are filed at any time from January through September (Internal Revenue Service, 2008). These differences mean that we do not expect the ACS, IRS and ASEC migration rates to match in any given year, but the change in migration rates over periods of several years should be comparable in the three datasets.

Figure 4(b) shows the interstate migration rates in the ASEC, the IRS data, and the ACS starting in 2005, the year that the ACS began to collect data from the entire United States.¹³ The ACS data are from the summary tables that the Census Bureau publishes on its American FactFinder web site¹⁴ and include both imputed and nonimputed data.¹⁵ The IRS rate is calculated from the migration tables on the IRS web site.¹⁶

¹²The ACS spends up to three months attempting to collect data from a given address, while the ASEC collects data only in a specific week, so if the address is vacant in a certain week but occupied in any of the next 12 weeks, the ACS will find a migrant where the ASEC would find a vacancy. The ACS and ASEC survey universes also differ in some ways.

¹³Before 2005, the Census Bureau's source and accuracy statements for the ACS say the survey is designed to represent only the counties where data were collected.

¹⁴"American FactFinder," U.S. Census Bureau, <http://factfinder.census.gov>, accessed Oct. 22, 2010.

¹⁵In the public-use data for the ACS, we have found that dropping imputed data has little effect on the estimated migration rate; however, results from the public-use file do not exactly match the summary tables because the public-use file includes only a subset of all ACS respondents.

¹⁶"SOI Tax Stats — Free Migration Data Downloads," Internal Revenue Service, <http://www.irs.gov/taxstats/article/0,,id=212718,00.html>, accessed Nov. 9, 2010.

The interstate migration rates in the IRS and ACS data actually rose between 2005 and 2006, contrary to the result in the ASEC. However, one-year fluctuations in the IRS and ACS data are difficult to interpret because of the imprecise timing of these datasets. Long-run changes are more instructive. From 2005 to 2009, the migration rate in the ACS fell 0.2 percentage point, from 2.48 percent to 2.28 percent. The IRS results almost exactly match the ACS for 2005 through 2008, the last year for which the IRS data are available. By contrast, in the ASEC the rate fell 1.02 percentage point from 2005 to 2009 when including imputed data, and 0.37 percentage point when we use only nonimputed data. The decline in interstate migration in the ACS and IRS data is thus even smaller than the decline in the ASEC and lends support to our argument that interstate migration has not decreased dramatically in recent years.

5. Policy and research implications

The precise timing of the ASEC makes it the most appropriate dataset for investigating how migration changed during the December 2007–June 2009 recession. Despite the media reports and concern among policymakers, there is no sign that interstate migration deviated meaningfully from its long-run trend. Recall that data for the 2007 survey year refer to migration between February–April 2006 and February–April 2007. Thus, the data points corresponding to the recession are those for the 2008, 2009, and 2010 survey years. Figure 4(a) shows that, in the nonimputed data that we view as most accurate, migration fell faster than trend in the 2007 survey year, *before* the recession.

As we have discussed, throwing out observations with missing or imputed data is not in general a good method. It happens to work well for the overall interstate migration rate because the current imputation procedure produces virtually identical rates for imputed and nonimputed data. The correspondence is not quite as close for some other rates, such as the rate of migration between counties in a state. There also is no guarantee that dropping imputed data would produce reliable results in an analysis of the relationship between migration and other variables. Further research is needed to find good methods for analyzing migration data in the ASEC despite the high rate of missing data and the change in the Census Bureau’s imputation procedures. In particular, it would be useful to investigate whether

researchers can construct reliable imputations using only the publicly available data, without access to the detailed geography and other variables in the Census Bureau’s internal files.

Understanding the causes of the long-run downward trend in migration rates is also an important topic for further research. Many factors may contribute to an individual’s decision about whether to migrate. These include the direct costs associated with moving houses, the loss of contact with family and friends, the benefits of access to a different labor market, and geographic differences in the availability of local amenities. Which of these factors is primarily responsible for the observed decline in migration will determine the appropriate policy response. For example, if migration is falling because the costs associated with moving are rising, then it may be difficult for workers to respond to economic shocks by moving to places with more productive opportunities. There might then be a need for government policies to, in the words of Assistant Treasury Secretary Krueger, “move the jobs — and create new jobs — in areas where the people are” (Fletcher, 2010, p. A1). On the other hand, high migration rates have often been associated with large economic differences between regions of the country, as in the case of the Great Migration of African Americans from the rural South to urban centers in the North and West (Curtis White et al., 2005; Hamilton, 1964). If migration rates are falling because the U.S. economy is becoming more homogeneous, the case for government intervention is less strong: Migration is low because it is unneeded, not because it is costly. Untangling the potential explanations for the long-run decline in migration is thus a valuable goal for future research. Our hope is that the correction for time-varying imputation procedures that we describe in this paper will help make such research possible, by allowing researchers to draw accurate inferences about interstate migration rates.

References

- Batini, Nicoletta, Oya Celasun, Thomas Dowling, Marcello Esteveão, Geoffrey Keim, Martin Sommer, and Evridiki Tsounta, 2010, “United States: Selected Issues Paper,” International Monetary Fund Country Report No. 10/248, available at <http://www.imf.org/external/pubs/ft/scr/2010/cr10248.pdf>.
- Brown, Charles, 1984, “Black-White Earnings Ratios since the Civil Rights Act of 1964: The Importance of Labor Market Dropouts,” *Quarterly Journal of Economics* 99(1), 31–44.
- Curtis White, Katherine J., Kyle Crowder, Stewart E. Tolnay, and Robert M. Adelman, “Race, Gender, and Marriage: Destination Selection during the Great Migration,” *Demography* 42(2), 215–241.
- Fletcher, Michael A., 2010, “Few in U.S. move for new jobs, fueling fear the economy might get stuck, too,” *Washington Post*, July 30, p. A1.
- Frey, William H., 2009, “The Great American Migration Slowdown: Regional and Metropolitan Dimensions,” Brookings Institution report.
- Hamilton, C. Horace, 1964, “The Negro Leaves the South,” *Demography* 1(1), 273–295.
- Internal Revenue Service, 2008, “Supplemental Documentation for Migration Data Products,” available at <http://www.irs.gov/pub/irs-soi/countmiguser0708.doc>.
- Koerber, Kin, 2007, “Comparison of ACS and ASEC Data on Geographic Mobility: 2004,” manuscript, U.S. Census Bureau, available at http://www.census.gov/acs/www/library/by_year/2007/.
- U.S. Census Bureau, 2006, “Design and Methodology: Current Population Survey,” Technical Paper 66, available at <http://www.census.gov/prod/2006pubs/tp-66.pdf>.