

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

Greg Kaplan · Sam Schulhofer-Wohl

Published online: 15 May 2012  
© Population Association of America 2012

**Abstract** We show that much of the recent reported decrease in interstate migration is a statistical artifact. Before 2006, the Census Bureau’s imputation procedure for dealing with missing data in the Current Population Survey inflated the estimated interstate migration rate. An undocumented change in the procedure corrected the problem starting in 2006, thus reducing the estimated migration rate. The change in imputation procedures explains 90% of the reported decrease in interstate migration between 2005 and 2006, and 42% 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. Contrary to popular belief, 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

## Introduction

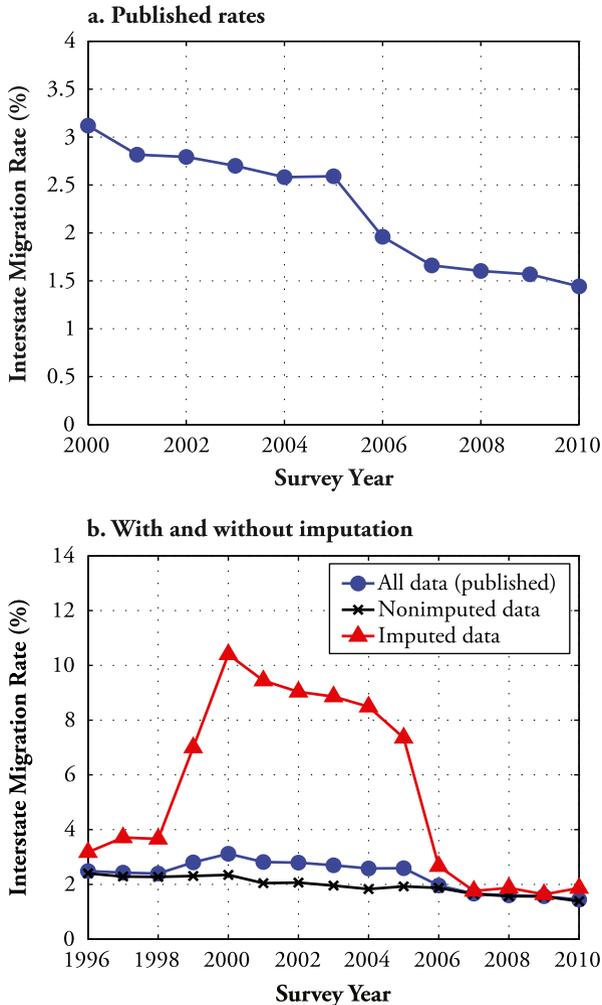
The low rate of migration between U.S. states has attracted much recent attention from demographers, policymakers, and the news media. Observers have often emphasized that in addition to the long-run downward trend in interstate migration over the past half-century, the migration rate fell sharply in 2006 from a relatively high plateau between 1999 and 2005. Panel a of Fig. 1 shows the annual interstate migration rate for the past decade, as calculated by the U.S. Census Bureau.<sup>1</sup> Taking

---

<sup>1</sup>Estimates were retrieved online (<http://www.census.gov/population/www/socdemo/migrate.html>).

G. Kaplan  
Department of Economics, University of Pennsylvania, Philadelphia, PA, USA

S. Schulhofer-Wohl (✉)  
Research Department, Federal Reserve Bank of Minneapolis, 90 Hennepin Ave., Minneapolis, MN  
55480-0291, USA  
e-mail: wohls@minneapolisfed.org



**Fig. 1** Rates of migration between states. Rate in panel a is the weighted percentage of respondents at least 1 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. In panel a, 2000–2009 survey years are from CPS ASEC tables (<http://www.census.gov/population/www/socdemo/migrate.html>), and 2010 survey year is from authors' calculations from CPS ASEC. Rate in panel b is the weighted percentage of respondents at least 1 year old who lived in a different state one year ago. Data in panel b are authors' calculations from CPS ASEC

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 (then) Assistant Treasury Secretary Alan B. Krueger, worry that low mobility will harm the nation’s recovery from the recession (Fletcher 2010:A1). Frey (2009:4), analyzing the Census Bureau data, called the decline in interstate migration “dramatic” and noted that “the 1.6 percent interstate migration rate for the past two years was half the value exhibited in 1999–2000.” And a recent International Monetary Fund report on

the U.S. economy argued that “slower inter-state migration, likely related to the housing crash,” was raising unemployment (Batini et al. 2010:5–6).

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

Panel b of Fig. 1 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 has published and the rate that recent research, policy, and media reports have discussed.<sup>2</sup>

The article proceeds as follows. The second section explains how the imputation procedures changed. Then we document the effect of the change on migration rates estimated from the ASEC. Following that, we argue that the nonimputed data provide the most accurate guide to recent trends in interstate migration and use independent data sets and an alternative imputation procedure to check the robustness of our results. We conclude by discussing how our findings relate to the policy discussion on migration in the current recession, and then we suggest avenues for further research.

### **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

---

<sup>2</sup> After 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 interstate migration rate. In response to this article, and after its circulation, the Census Bureau added a note on its website (<http://www.census.gov/population/www/socdemo/CPSnote.pdf>) that explains the change in imputation procedure and cautions users about its effects on measured migration rates. The change is also mentioned in a footnote in an unpublished Census Bureau working paper that compares the ASEC with the American Community Survey (Koerber 2007:14).

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.<sup>3</sup> Unlike data on tax return mobility from the Internal Revenue Service (IRS), the ASEC mobility data are part of a data set 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.<sup>4</sup> 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 significant item nonresponse: cases in which 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 for 10% to 12% of ASEC respondents. It would be inappropriate to calculate migration rates using only the data from people who answer the questions because those who do not answer may differ along both observable and unobservable dimensions from those who do answer. In principle, one could calculate bounds on the migration rate by assuming alternatively that all nonrespondents are migrants or that none are.<sup>5</sup> However, because nonresponse is so common, the bounds would be so wide as to be almost useless.

To account for observable differences between respondents and nonrespondents, the Census Bureau uses an imputation procedure known as hot deck allocation. The technical documentation for the CPS (U.S. Census Bureau 2006: chap. 9) describes the method in detail. Hot deck allocation works by defining a set of cells based on observable characteristics,  $X$ . Missing values are filled in for one respondent at a time. If a respondent did not answer a question, an imputed answer is generated by copying the answer of the most recently processed respondent who falls in the same  $X$  cell and who did answer the question. The person whose missing answer is filled in is referred to as a “recipient”; the person whose answer is used is a “donor.” If, conditional on  $X$ , there are no differences between nonrespondents and respondents in unobservable characteristics that affect the variable being imputed, this method will generate unbiased estimates of the mean of the variable that is being imputed. However, this method will still lead to biased estimates if unobservable differences between respondents and nonrespondents remain after conditioning on  $X$ .

Which characteristics are included in  $X$  depends on what variable is being imputed. According to Koerber (2007), for imputing mobility variables, the 2004 ASEC used

<sup>3</sup> The 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.

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

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

cells based on  $X = \{\text{census division of current residence, race, housing tenure, age, armed forces status}\}$ . Imputed migration rates will differ from nonimputed rates only if the distribution of these characteristics differs among respondents and nonrespondents. Thus, if the characteristics included in  $X$  are changed, the resulting imputations will change. The change in processing order that we describe herein implicitly changed the effective set  $X$  used by the Census Bureau, which in turn changed the imputed migration rates.

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. After 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 his or 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. Because 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 recipient moved, it usually imputes a local move. Before 2006, the order of processing was geographic but within particular samples.<sup>6</sup> (The ASEC consists of several samples, not one; see U.S. Census Bureau 2006: chap. 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 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—say, Minneapolis—and the donor's location one year ago was probably near Minneapolis, not Philadelphia. The pre-2006 less-geographic 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 instead because the nonrespondent is similar to a respondent who moved a short distance in Minnesota.

In formal terms, the change in sort order changed the effective amount of geographic information included in the conditioning variables  $X$ . The new sort

<sup>6</sup> Personal communication 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).

order is likely to result in better imputations because current location is a good predictor of previous location: the more information about current location is included in  $X$ , the less selection on unobservables in previous location there will be. However, our concern here is with measuring migration trends over time, and for that purpose, it is important to hold constant the amount of selection on unobservables over time. Any *change* from year to year in the sort order or conditioning variables—regardless of which sort order or conditioning set is better—will change the imputed data, but such a change should not be confused with a change in actual migration behavior.

An important unresolved puzzle remains: why the imputed interstate migration rate for nonrespondents jumped sharply in 1999 and again in 2000, while the migration rate among respondents did not. Panel b of Fig. 1 shows an increase of more than 6 percentage points over this two-year period in the imputed data. Despite our best efforts, including a close examination of the technical documentation for the CPS (U.S. Census Bureau 2000, 2002, 2006), we have not been able to uncover the reasons for this increase. 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—from Fortran to SAS. However, staff also told us that the change in computer languages should not have affected the results.<sup>7</sup> Without access to nonpublic elements of the data and the Census Bureau's imputation programs, it may be impossible for outside researchers to determine what happened in 1999 and 2000. Because it is likely that the sharp jump in imputed migration rates in these years is the root source of the subsequent problems, we urge the Census Bureau to investigate this issue further.

### 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 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.<sup>8</sup> We consider a respondent to have imputed data only if the data came from hot deck

<sup>7</sup> In Kaplan and Schulhofer-Wohl (2011), we discussed and ruled out alternative explanations for the 1999 and 2000 changes, including changes in the CPS weights associated with the 2000 census and the expansion of the ASEC sample in 2001.

<sup>8</sup> In 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.

allocation. We categorize respondents as nonimputed if their migration data were inferred from household members' survey responses.<sup>9</sup>

In panel b of Fig. 1, we demonstrate 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.<sup>10</sup> Panel a of Fig. 2 shows that the change in imputation procedures had the opposite effect on within-county migration rates as on interstate migration rates: 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. Panel b of Fig. 2 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.

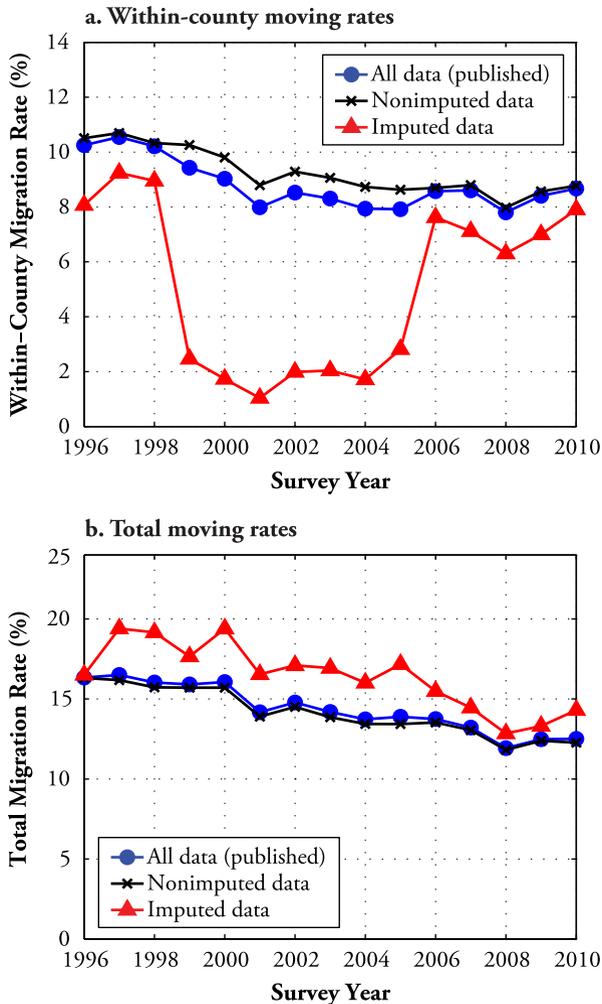
The comparisons of migration rates in imputed and nonimputed data do not, of course, prove that the pre-2006 imputation procedure led to an artificially inflated interstate migration rate and a reduced within-county rate. Ideally, we would show the effect of the change in imputation procedures by redoing the pre-2006 imputations using the sort order in effect since 2006. Although even the new procedure does not necessarily generate accurate *levels* of migration rates because there could be selection on unobservables conditional on the variables used to define imputation cells, holding the imputation procedure constant would highlight the effect on the migration *trend* of the implicit change in the conditioning variables induced by the change in sort order. Provided that the degree of selection on unobservables is not changing over time, a time-invariant imputation procedure would produce an accurate estimate of the migration trend. 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) indicated

---

<sup>9</sup> Specifically, 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 the person is coded as a migrant and state of residence one year ago was hot deck allocated; if 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% 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 because there is no information to indicate that their data came from hot deck allocation. However, the migration rates in unlinkable observations follow the same pattern as the rates in imputed data, so some researchers may prefer to treat unlinkable observations as imputed. Classifying unlinkable observations as hot deck allocated does not significantly change our findings.

<sup>10</sup> The 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.



**Fig. 2** Moving rates. Rate in panel a is the weighted percentage of respondents at least 1 year old who lived in a different home in the same county one year ago. Rate in panel b is the weighted percentage of respondents at least 1 year old who lived in a different home one year ago. Data in both panels are authors' calculations from CPS ASEC

that the pre-2006 sort order raised the interstate migration rate relative to a geographic sort used in the ACS.

### What Is the True Trend in Interstate Migration?

We agree with the Census Bureau that the geographic sort 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. (Alternatively, as we illustrate shortly, one could reduce the sort order's impact by imputing the type of move instead of inferring the type of move

from an imputed location one year ago.) The pre-2006 data likely overstate the rate of interstate migration and understate the rate of local migration.

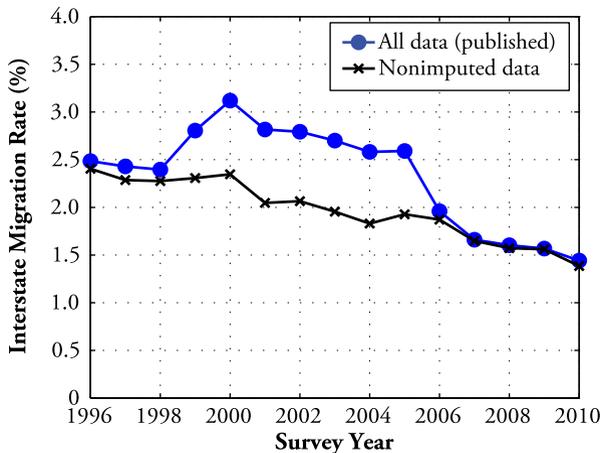
The change in imputation procedures in 2006 means that comparisons of pre-2006 and post-2006 data do not accurately measure trends in interstate migration. However, because there may have been good reasons for using different sort orders to obtain better imputations of other variables in past surveys, we do not think it would necessarily be desirable for the Census Bureau to recalculate the pre-2006 imputations using the current procedure.<sup>11</sup> Fortunately, there is a simple way to obtain accurate estimates of the trend in interstate migration rates over the last decade. Since 2006, the interstate migration rate including imputed data has been virtually identical to the rate using only nonimputed data. Figure 3 illustrates this point by reproducing Fig. 1, panel 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 can focus on the rate in nonimputed data, and doing so will remove the fluctuations induced by changes in the imputation procedure.

Figure 3 shows that after 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% in the 2000 survey, fell to 2.59% by the 2005 survey, plummeted to 1.96% in the 2006 survey, and is now down to 1.44%. Without imputations, the rate was 2.35% in the 2000 survey, 1.93% in the 2005 survey, 1.87% in the 2006 survey, and 1.38% 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 carry out a partial check on how much the missing data matter for the trend in interstate migration by constructing our own imputations using a procedure that does not vary over time. We call this a partial check because, although our procedure corrects for observable differences between nonrespondents and respondents, it—like all imputation methods—does not correct for selection on unobservables. But this partial check is valuable because it illustrates what the trend in estimated interstate migration rates might have been if the Census Bureau had kept its imputation procedures constant from year to year.

We follow the Census Bureau in dividing people into cells based on census division of current residence, race, housing tenure, age, and armed forces status. However, instead of imputing migration status and location of previous residence and then calculating how far the nonrespondent moved, we directly impute an indicator variable for whether the nonrespondent moved between states. In particular, we randomly designate each nonrespondent in a cell as an interstate migrant with probability equal to the weighted fraction of respondents in the cell who are interstate migrants. We use multiple imputation (Rubin 1987) to account for uncertainty in the imputed results. For details, see Kaplan and Schulhofer-Wohl (2011).

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



**Fig. 3** Rates of migration between states: detail. Rate is the weighted percentage of respondents at least 1 year old who lived in a different state one year ago. Data are authors' calculations from CPS ASEC

Our imputation procedure is equivalent to hot deck imputation if people are sorted in random rather than geographic order and if the interstate migration indicator is imputed directly instead of calculated as a function of imputed previous location. We can avoid geographic sorting because although location of current residence is highly informative about location of previous residence, it is less informative about migration status *per se*, and we are only imputing migration status. An advantage of our procedure is that because it does not sort respondents geographically, it does not require detailed address information that the Census Bureau withholds from public-use data files.

Table 1 shows the results. The interstate migration rate that we calculate using the completed data—the actual data of respondents and the imputed data of nonrespondents—is virtually identical to the migration rate using only nonimputed data in each year from 1996 to 2010, supporting our argument that the nonimputed data provide a good guide to the trend in interstate migration. The Census Bureau's published rate including imputed data is slightly higher than our rate, perhaps because the bureau's two-step method of imputing location last year and then calculating the distance of the move can impute some spurious interstate moves—although these should be rare when the data are sorted geographically.

The close correspondence between our imputations and the nonimputed data reflects the fact that there are few large differences between respondents and nonrespondents in the conditioning variables  $X$  on which we (and the Census Bureau) base the imputations, and what differences exist largely cancel out. We show in Kaplan and Schulhofer-Wohl (2011) that nonrespondents are, on average, older than respondents, less likely to be white, and less likely to be homeowners. Imputation corrects for all these observable differences. However, the corrections have offsetting effects: among respondents, whose data are used to construct the imputation, older people and nonwhites have lower migration rates, while nonhomeowners have higher migration rates.

**Table 1** Alternative imputation of interstate migration rates in the CPS

Survey Year	All Data	Nonimputed Data	Multiple Imputation Results	
			Estimated Rate	SE
1996	2.4851	2.4042	2.3925	0.0510
1997	2.4293	2.2866	2.2811	0.0496
1998	2.3964	2.2757	2.2766	0.0490
1999	2.8051	2.3070	2.3123	0.0509
2000	3.1193	2.3473	2.3440	0.0502
2001	2.8164	2.0472	2.0528	0.0486
2002	2.7931	2.0651	2.0706	0.0406
2003	2.6995	1.9547	1.9673	0.0394
2004	2.5816	1.8298	1.8351	0.0380
2005	2.5912	1.9277	1.9378	0.0423
2006	1.9599	1.8718	1.8943	0.0401
2007	1.6610	1.6480	1.6551	0.0377
2008	1.6034	1.5730	1.5820	0.0372
2009	1.5681	1.5602	1.5740	0.0362
2010	1.4416	1.3848	1.3907	0.0340

*Note:* Rate is the weighted percentage of respondents at least 1 year old who lived in a different state one year ago.

*Source:* Authors' calculations from CPS ASEC.

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 with 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) noted that the ACS tends to find more migrants than the ASEC because of differences in survey procedures.<sup>12</sup> Meanwhile, the IRS data cover only those people who file income tax returns, which means that 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 data sets.

Figure 4 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

<sup>12</sup> The ACS spends up to three months attempting to collect data from a given address, while the ASEC collects data only in a specific week. Thus, 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.

United States.<sup>13</sup> We calculate the ACS rates using the IPUMS-USA version of the ACS public-use file (Ruggles et al. 2010).<sup>14</sup> The calculation includes both imputed and nonimputed data; dropping imputed data changes the results by only a few hundredths of a percentage point in each year.<sup>15</sup> We calculate the IRS rate from the migration tables available online from the IRS.<sup>16</sup>

The interstate migration rates in the IRS and ACS data actually rose between 2005 and 2006, contrary to the result in the ASEC. However, the rise in the ACS data is due to the addition of group-quarters residents to the sample in 2006. When we remove group-quarters residents, the ACS rate drops slightly from 2005 to 2006. Also, one-year fluctuations in the IRS and ACS data are difficult to interpret because of the imprecise timing of these data sets. Long-run changes are more instructive. From 2005 to 2009, the migration rate in the ACS fell 0.4 percentage point, from 2.48% to 2.08%, after we exclude group-quarters residents. The IRS rate fell somewhat less than the ACS rate 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 points from 2005 to 2009 when we include 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 the same as or smaller than the decline in the ASEC.

## Policy and Research Implications

The precise timing of the ASEC makes it the most appropriate data set 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 3 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.<sup>17</sup>

As we have discussed, throwing out observations with missing or imputed data is not generally recommended. It happens to be a reasonable choice when estimating the interstate migration rate because the observable differences between respondents and

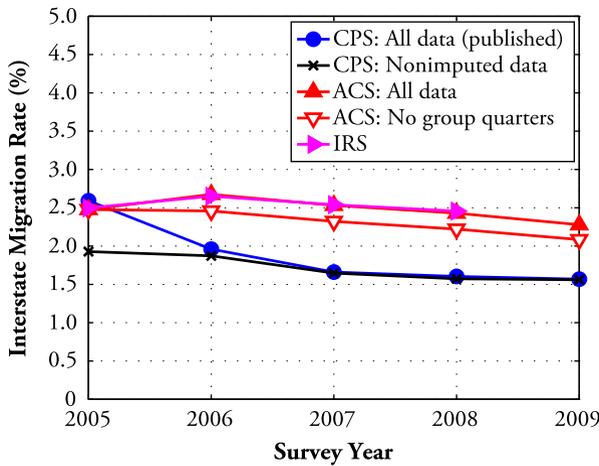
<sup>13</sup> Before 2005, the Census Bureau's source and accuracy statements for the ACS say the survey is designed to represent only the counties where data were collected.

<sup>14</sup> The summary tables that the Census Bureau publishes on its American FactFinder web site (<http://factfinder2.census.gov>), which are based on a somewhat larger data file, show a virtually identical overall interstate migration rate but do not allow us to distinguish imputed from nonimputed data or group-quarters residents from others.

<sup>15</sup> Koerber (2007) reported that the ACS uses a hot deck allocation procedure in which donors and recipients always have the same current state of residence. Therefore, the ACS procedure does not generate the spurious interstate moves that the pre-2006 ASEC method produced when donors came from different states than recipients.

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

<sup>17</sup> Although we hesitate to overemphasize small fluctuations in the data, we conjecture that the slight increase relative to trend in 2006 and drop relative to trend in 2007 may be related either to fluctuations in the housing market or to migration induced by Hurricane Katrina in August 2005.



**Fig. 4** Comparison of data sources. CPS and ACS rates are weighted percentages of respondents at least 1 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. Data are authors’ calculations from CPS ASEC, IRS Statistics of Income, and ACS public-use data

nonrespondents have little net effect on interstate migration. We also show in this article that a time-invariant imputation procedure yields virtually identical results to dropping nonrespondents. However, the choice of an appropriate imputation method depends on the research context, so users should evaluate any imputation method to ensure that it is suitable for their particular purposes. To facilitate such evaluation, we encourage the Census Bureau to release full documentation about editing and allocation procedures for all variables on all surveys, except when releasing the documentation would compromise respondents’ confidentiality. In addition, further research is needed on whether unobservable differences between respondents and nonrespondents may bias estimates of migration rates—a problem that imputation cannot correct.

Many factors may contribute to decisions about whether to migrate; which factors are primarily responsible for the observed decline in migration will determine the appropriate policy response. For example, if migration is falling because the costs of moving are rising, then it may be difficult for workers to respond to economic shocks by moving to places with more opportunities. There might then be a need for government policies to, in the words of (then) Assistant Treasury Secretary Krueger, “move the jobs—and create new jobs—in areas where the people are” (Fletcher 2010: 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 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 research. We hope the correction for time-varying

imputation procedures that we describe in this article will help make such research possible by allowing researchers to draw accurate inferences about interstate migration rates.

**Acknowledgments** We thank Andrew Gelman, Ellen McGrattan, Robert Moffitt, an anonymous commenter, and the Editor and referees of this journal for helpful suggestions; Joan Gieseke for editorial assistance; and Xun Liu for excellent research 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.

## References

- Batini, N., Celasun, O., Dowling, T., Estevão, M., Keim, G., Sommer, M., Tsounta, E. (2010). *United States: Selected issues paper* (IMF Country Report No. 10/248). Washington, DC: International Monetary Fund. Retrieved from <http://www.imf.org/external/pubs/ft/scr/2010/cr10248.pdf>
- Brown, C. (1984). Black-white earnings ratios since the Civil Rights Act of 1964: The importance of labor market dropouts. *Quarterly Journal of Economics*, 99, 31–44.
- Curtis White, K. J., Crowder, K., Tolnay, S. E., & Adelman, R. M. (2005). Race, gender, and marriage: Destination selection during the Great Migration. *Demography*, 42, 215–241.
- Fletcher, M. A. (2010, July 30). Few in U.S. move for new jobs, fueling fear the economy might get stuck, too. *Washington Post*, p. A1.
- Frey, W. H. (2009). *The Great American Migration slowdown: Regional and metropolitan dimensions* (Report). Washington, DC: Brookings Institution.
- Hamilton, C. H. (1964). The Negro Leaves the South. *Demography*, 1, 273–295.
- Internal Revenue Service. (2008). Supplemental documentation for migration data products. Retrieved from <http://www.irs.gov/pub/irs-soi/countmiguser0708.doc>
- Kaplan, G., & Schulhofer-Wohl, S. (2011). *Interstate migration has fallen less than you think: Consequences of hot deck imputation in the Current Population Survey* (Federal Reserve Bank of Minneapolis Working Paper 681).
- Koerber, K. (2007). *Comparison of ACS and ASEC Data on Geographic Mobility: 2004* (Report). Washington, DC: U.S. Census Bureau. Retrieved from [http://www.census.gov/acs/www/library/by\\_year/2007/](http://www.census.gov/acs/www/library/by_year/2007/)
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: John Wiley & Sons.
- Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B., & Sobek, S. (2010). *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota.
- U.S. Census Bureau. (2000). *Current Population Survey: Design and methodology* (Technical Paper 63). Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/prod/2000pubs/tp63.pdf>
- U.S. Census Bureau. (2002). *Current Population Survey: Design and methodology* (Technical Paper 63RV). Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/prod/2002pubs/tp63rv.pdf>
- U.S. Census Bureau. (2006). *Current Population Survey: Design and methodology* (Technical Paper 66). Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/prod/2006pubs/tp-66.pdf>