SOI Migration Data: A New Approach

Methodological Improvements for SOI's United States Population Migration Data, Calendar Years 2011–2012

by Kevin Pierce

The IRS Statistics of Income Division (SOI), in collaboration with the U.S. Census Bureau, has released migration data for the United States for several decades. These data are an important source of information detailing the movement of individuals from one location to another. SOI bases these data on the year-to-year address changes reported on individual income tax returns filed with the IRS during two consecutive calendar years.

From the migration data's inception through Calendar Years 2009–2010, the Census Bureau produced the data for SOI. This process all changed beginning with data for 2011–2012 when SOI assumed the responsibility for the migration tabulations and introduced a number of enhancements intended to improve the data's overall quality. Furthermore, the new approach provided an additional series of information. This paper discusses those improvements and highlights some of the differences between the migration data for 2011–2012 and previous versions.[1] To date, SOI has made the following three major improvements:

- Migration data are now based on a full year of data, as opposed to a partial year of data.
- Overall, the improved year-to-year return matching has increased the number of matched records by 5 percent and the number of high-income returns by approximately 25 percent.
- New tabulations show migration flows at the State level, by size of adjusted gross income (AGI) and age of primary taxpayer.

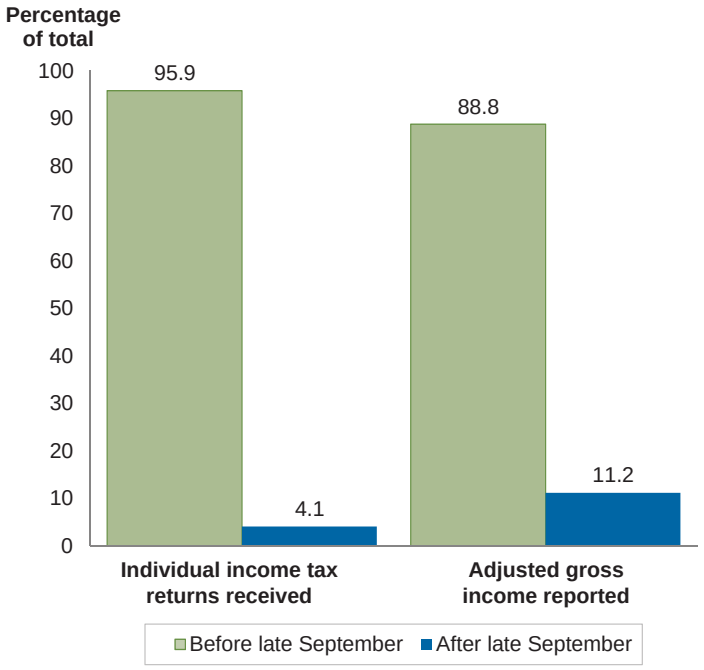
Full-Year Data

A major improvement made to the data was the switch from partial- to full-year data. Due to their internal production deadlines, migration data produced by Census for SOI only included individual income tax returns filed before the end of September. Since SOI is not bound by those same deadlines, it can construct the migration tabulations using all individual income tax returns received during the full calendar year (January 1 to December 31).[2]

For 2012, some 4 percent of the returns filed with the IRS were received after the September cutoff (Figure A). Also, the total adjusted gross income (AGI) reported for returns received after September was equal to 11 percent of the total AGI for 2012. To the extent that returns filed after September are somewhat different from those filed before September (more complex and with higher income, for instance), including them in the migration data removes a potential bias. Another way of stating this is that by including the returns received after late September, the new SOI migration data represent the migration patterns for the full filing population of current tax year returns.

Figure A

Percentages of Individual Income Tax Returns Received and Adjusted Gross Income Reported Before and After Late September, Calendar Year 2012



Source: IRS, Statistics of Income Division, Individual Master File, July 2015.

Improved Year-to-Year Return Matching

Prior to the 2011–2012 migration data, Census matched returns based on the primary filer's taxpayer identification number (TIN) only. One drawback to this is that the filing position for an individual can change over time. For instance, individuals may be listed as a dependent on a parent or guardian's return in Year 1, but file their own return as a primary filer in Year 2. Similarly, an individual listed as a secondary filer on, for example, a married-filing-joint return in Year 1 could become a single filer

[1] For a full description of the 2011–2012 migration data, see the Migration Data Users Guide available at http://www.irs.gov/uac/SOI-Tax-Stats-Migration-Data. For a description of the previous migration data, see Gross, Emily, "U.S. Population Migration Data: Strengths and Limitations," available on this Webpage.

[2] The new migration data include all individual income tax returns processed from January 1 to December 31 for the current tax year. A small percentage of the returns that the IRS receives (usually around 3 percent) are for prior tax years. For matching purposes, SOI excludes these returns from the migration data.

(primary taxpayer) in Year 2 due to divorce or death. Finally, some married couples switch their positions each year, the primary becoming the secondary and vice versa. Under the previous methodology, records from the migration data that had experienced such changes in taxpayers' filing positions were excluded.

To address this limitation, the matching process under the new methodology uses the TINs of the primary, secondary, and dependent filers. Using this method, approximately 5 percent of the total matched records occur through additional matching beyond the primary-to-primary matching (Figure B). Dependent filers (in Year 1) who matched to primary filers (in Year 2) comprised the largest source of these additional records (around 3 percent).

Despite these improvements, a number of records can still be excluded from the final migration data. A nonmatching return can occur if a TIN is recorded incorrectly between the two years; if a taxpayer switches from a temporary TIN to a permanent Social Security Number (SSN); or if a taxpayer filed a return in one year, but did not file a return timely in another year.[3] Overall, the additional merging of secondary and dependent filers increases the precision of the migration data by including a wider segment of the filing population.

Figure B

SOI Migration Data: New Matching Process for Individual Income Tax Returns

Year 1	Year 2	Percentage of the total matched returns
Primary filer	Primary filer	94.6
Primary filer	Secondary filer	0.8
Secondary filer	Primary filer	1.7
Secondary filer	Secondary filer	less than 0.1
Dependent filers	Primary filer	2.8
Dependent filers	Secondary filer	less than 0.1

IRS, Statistics of Income Division, Individual Master File, July 2015.

New Tabulations—The Gross Migration File

The last major improvement introduces a new tabulation that shows aggregate migration flows at the State level, by size of adjusted gross income (AGI) and age of the primary taxpayer. The Gross Migration File is a summary of the migration flows for each State, plus the District of Columbia, that shows the total number of matched returns, nonmigrant returns, outflow returns, inflow returns, and same-State returns.

Comparisons of the Old and New Methodologies

In general, SOI has made every effort to preserve the comparability and continuity of the new migration data with the previous years of data. The new migration data include the same set of files: State-to-State outflows and inflows and county-to-county outflows and inflows; as well as the same codes and definitions used in previous years. However, due to the methodological changes between the two, the data are not directly comparable.

Matched Returns by Size of Adjusted Gross Income

For the 2011–2012 migration data, SOI created a test file that mimicked the same procedures the Census Bureau employed to create the previous migration data.[4] This file resulted in a total match of 111.5 million returns (Figure C). Using the updated methodology, the new version of migration data had 116.8 million returns, a 4.7-percent increase, or 5.3 million returns from the previous version. Returns in the $200,000-or-more AGI category experienced the greatest increase, rising about 25 percent, from 3.6 million to 4.5 million matched individual income tax returns.

Figure C

SOI Migration Data: Total Number of Matched Individual Income Tax Returns, by Size of Adjusted Gross Income, Calendar Years 2011–2012

Size of adjusted gross income	Number of matched returns		
	Previous methodology	New methodology	Percentage change
Total	**111,505,981**	**116,764,589**	**4.7**
$1 under $10,000	10,380,065	10,846,031	4.5
$10,000 under $25,000	26,886,965	27,607,234	2.7
$25,000 under $50,000	29,804,233	30,755,044	3.2
$50,000 under $75,000	17,007,772	17,717,602	4.2
$75,000 under $100,000	10,723,397	11,253,887	4.9
$100,000 under $200,000	13,109,345	14,107,846	7.6
$200,000 or more	3,594,204	4,476,945	24.6

IRS, Statistics of Income Division, Individual Master File, July 2015.

Additional Matched Returns

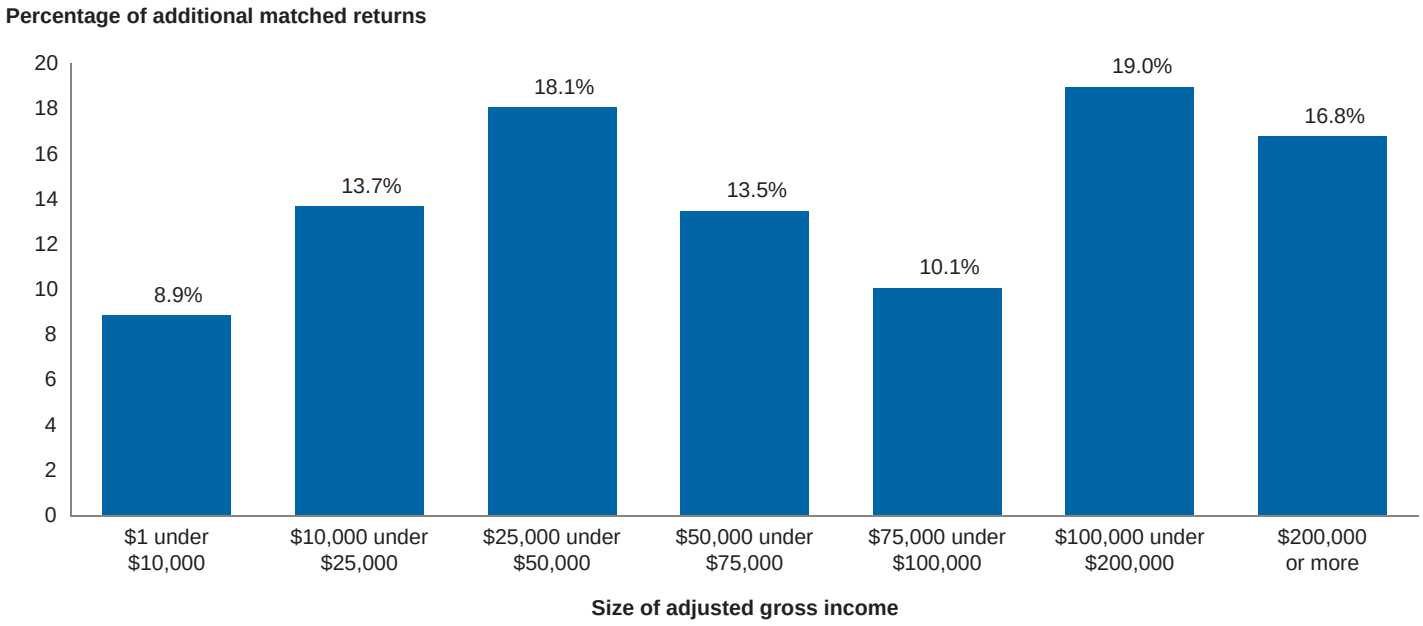
Of the 5.3 million records added to the migration data, over a third (35.8 percent) came from the top two income categories (Figure D). Matched returns in the $100,000-under-$200,000 category comprised 19 percent of the additional records, while returns in the $200,000-or-more category made up 16.8 percent.

[3] Individuals can apply to the IRS for an Individual Taxpayer Identification Number (ITIN) to file a valid U.S. Federal income tax return. An ITIN is a special 9-digit tax processing number, beginning with the number "9." There are instances where some individuals will receive a valid Social Security Number (SSN) in place of their ITIN and must file their individual return using the SSN. Excluded from the data are returns that switch between an ITIN and an SSN between migration years because of the nonmatching TINs.

[4] SOI created a test file to compare the previous methodology, data produced for Calendar Years 2009–2010 and earlier, to the new methodology, data produced for Calendar Years 2011–2012 and beyond. To avoid complementary disclosure, SOI is unable to release this file to the public.

Figure D

SOI Migration Data: Percentage of Additional Matched Individual Income Tax Returns, by Size of Adjusted Gross Income, Calendar Years 2011–2012



IRS, Statistics of Income Division, Individual Master File, July 2015.

As discussed earlier, by including full-year data, the new migration data represent to a greater degree the movement or nonmovement of high-income individuals and the income associated with those returns.

Distribution of Returns by Size of Adjusted Gross Income

In spite of the additional records, the distribution of returns by the size of adjusted gross income is fairly consistent between the test dataset and the new one (Figure E). The two largest AGI categories, in terms of the percentage of matched returns, decreased slightly in the new migration data relative to the data using the previous methodology. Matched returns in the $10,000-under-$25,000 category declined slightly from 24.1 percent to 23.6 percent. Returns in the $25,000-under-$50,000 category declined from 26.7 percent to 26.3 percent. In contrast, matched returns in the $100,000-under-$200,000 and $200,000-or-more categories experienced slight increases.

Net Migration Rates

An alternate method of comparing the previous methodology with the new methodology is by examining the net migration rate between the two. SOI calculates the net migration rates by subtracting the number of out-migrant returns from the number of in-migrant returns and dividing this net amount by the sum of the nonmigrant returns plus the out-migrant returns. Despite the differences between the old and new methodologies, a majority of States (86.2 percent) had a percentage difference of less than 5 percent (Figure F). A handful of States had modest differences between the old and new migration rates, and one State had a percentage difference that was greater than 10 percent.

Wyoming had the largest percentage difference between the calculated net-migration rate of the previous methodology and the new version (Figure G). Wyoming's percentage difference of 27.6 percent was nearly three times as high as the second largest percentage difference of 9.4 percent for South Dakota.

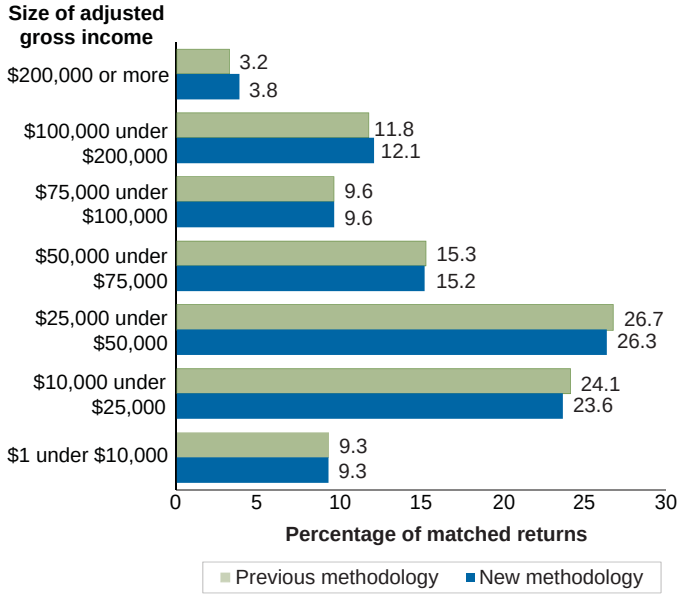
Additional Information

SOI's migration data present migration patterns by State or by county for the entire United States and are available for inflows—the number of new residents who moved to a State or county and where they migrated from, and outflows—the number of residents leaving a State or county and where they went. The data are available at http://www.irs.gov/uac/SOI-Tax-Stats-Migration-Data.

Kevin Pierce is an economist with the Individual Returns Analysis Section. This article was prepared under the direction of Michael Weber, Chief, Individual Returns Analysis Section.

Figure E

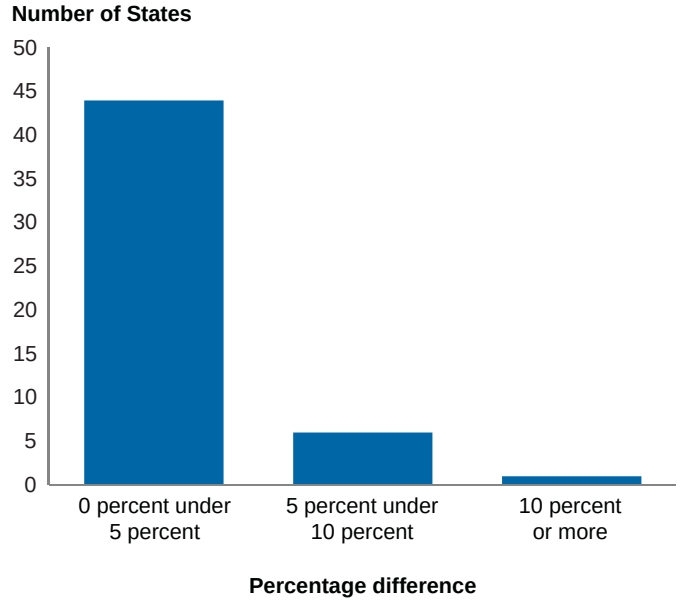
SOI Migration Data: Distribution of Matched Individual Income Tax Returns, by Size of Adjusted Gross Income, Calendar Years 2011–2012



IRS, Statistics of Income Division, Individual Master File, July 2015.

Figure F

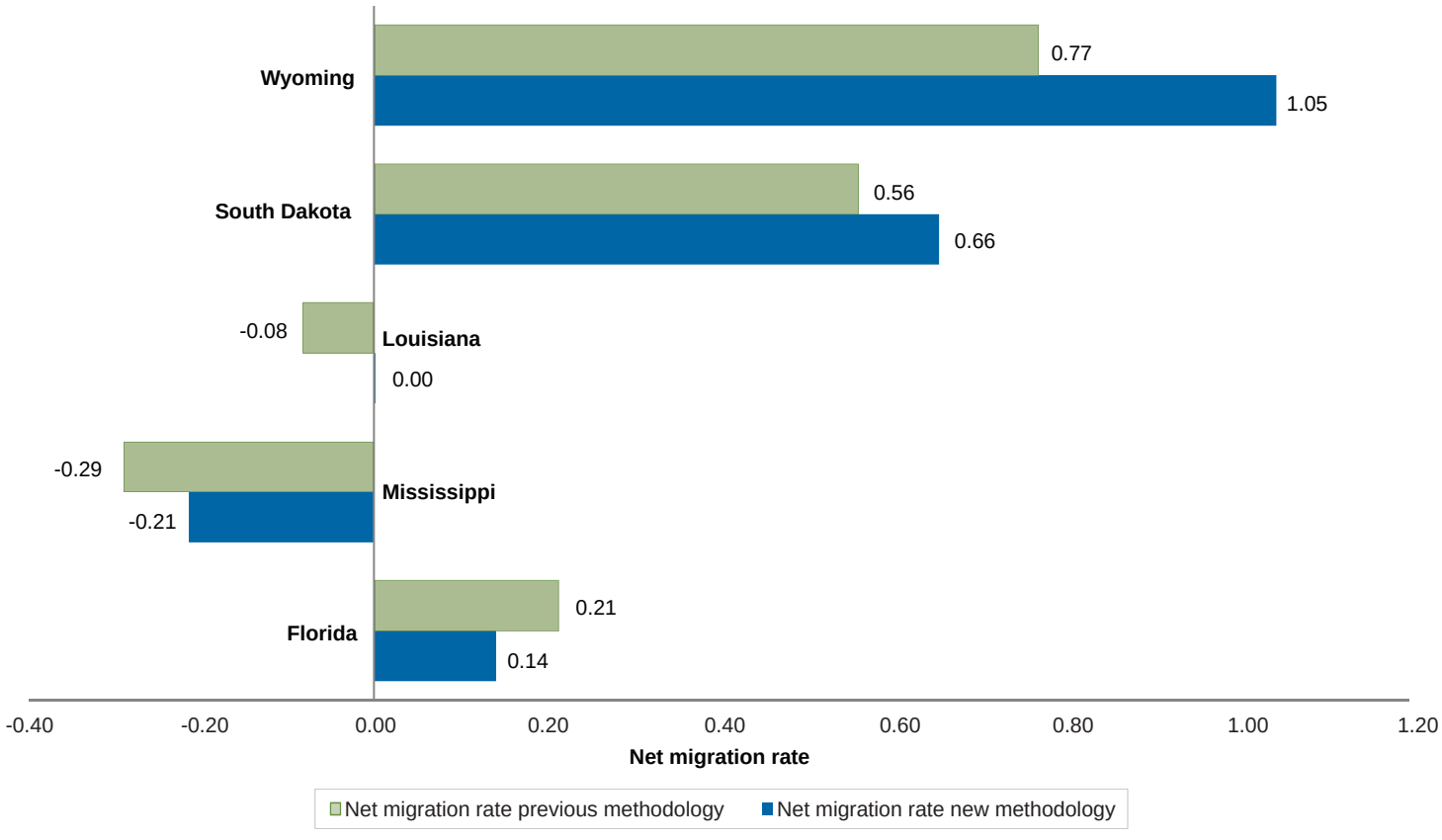
SOI Migration Data: Number of States, by Percentage Difference Between the Previous and New Methodologies' Net Migration Rates, Calendar Year 2011–2012



IRS. Statistics of Income Division. Individual Master File. Julv 2015.

Figure G

SOI Migration Data: Top 5 States with the Largest Net Migration Rate Differential, Calendar Years 2011–2012



IRS, Statistics of Income Division, Individual Master File, July 2015.