Internal Revenue Service Area-To-Area Migration Data: Strengths, Limitations, and Current Trends

Emily Gross, Internal Revenue Service

The mobility of Americans has long been a subject of interest for demographers, scholars, and the media. Just a few decades ago, the ultimate success story in this country was home ownership and staying in one neighborhood for all of adulthood. Currently, people and families move many times during their adult lives, with the peak moving years being between 20-24 years of age.[1] To where are these people moving, and from where did they originate? One of the few accurate sources of area-to-area migration data in the United States comes from the Statistics of Income Division (SOI) of the Internal Revenue Service (IRS), which maintains records of all individual income tax forms filed in each year.

This paper will highlight the data IRS has on taxpayer migration, particularly the county-to-county migration data created by U.S. Bureau of the Census analysts using IRS data. First, the paper will discuss the IRS Individual Master File from which these datasets are derived. Then, it will cover how the Census Bureau reviews the file and runs it through a geocoding program. Next, the paper will cover how the dataset returns to the IRS for disclosure proofing and how the data are marketed. The data themselves will be discussed, highlighting strengths and limitations. Finally, some current trends in migration will be examined.

▶ Statistics of Income (SOI) Division and the Data Source

The Statistics of Income program began in 1916, when Congress passed a revenue act that included a provision requiring the annual compilation of statistics with respect to the operation of the tax law. This requirement has reappeared in each major rewrite of the tax law since then and is currently included as section 6108 of the Internal Revenue Code of 1986.

Besides annual SOI publications, based on individual and corporate income tax returns, other data are also published in the quarterly *Statistics of Income Bulletin*. The *Bulletin* includes studies on sole proprietorships, partnerships, tax-exempt organizations, estate tax returns, and estimates of personal wealth, as well as studies on "international" tax returns. Most of the SOI publications are available on the "tax stats" portion of the IRS Web site (www.irs.gov), which contains over 3,900 files related to tax statistics.

From time to time, SOI undertakes special reimbursable studies for Government and private users. One customer, the Census Bureau (which is allowed access to tax return data under the Internal Revenue Code but must be able to justify the data items it receives as needed for its own statistical programs) pays IRS for annual data on every entity on the IRS Individual Master File (IMF). (The IRS Master File includes administrative records for every Form 1040, 1040A, and 1040EZ.) The tax and income items that Census receives from the IMF include:

- Tax Filing Units (the filer and spouse of filer, plus all exemptions represented on the forms)
- Mailing address
- Age classification (the filer is classified as "under age 65" if he or she did not mark the age 65+ checkoff box)
- Income data: wages and salaries, interest income, dividend income, gross rents, and royalties
- Adjusted gross income (includes all taxable income, less adjustments to income)
- Total income (a special definition which most closely approximates the Census Bureau's definition of total income).

The Master File data that Census receives were based on all returns filed by late September of the filing year. This extract is believed to include 95 percent to 98

percent of the individual filing population. The individuals covered by the returns include the filer and the spouse of the filer, as well as any exemptions claimed on the tax return. The Tax Year 2002 file, the most recent data available, contained about 130.5 million returns.[2]

In addition to using these data for their population estimates, Census also uses them to produce area-to-area migration data for SOI. The tax and income data included in the migration data are Number of Returns, Number of Exemptions, Aggregate Adjusted Gross Income (AGI), and Median AGI.

▶ Census Bureau Processing

In accordance with the agreement mentioned above between the IRS and Census Bureau, the 1040 Individual Master File dataset is provided annually to the Planning, Research, and Evaluation Department at Census. Both the Social Security Number (SSN) and the taxpayer name are stripped from each return. In their place, a special identification number called a Protective Identification Key (PIK) is assigned to each return.

To further prepare the data for its own purposes, as well as to prepare the migration files, the Census Bureau geocodes the IMF data. Geocoding involves assigning a set of codes to each return that represent the residence of the filer. These codes are assigned from the United States Post Office (USPO) ZIP/Sector-to-County Cross Reference (CCRS), which is generally reflected in the "ZIP plus 4" codes. The "plus 4" codes have two characters each--a sector code and a segment code. According to USPO guidelines, each sector code must identify one county only. This is the key to how Census is able to geocode each return by county of origin. From the combination of ZIP sector codes and mailing State code for each individual return, Census is able to assign each record with a State/county code from the CCRS. To prepare the migration data, Census must use 2 consecutive filing years of IMF data. For each set of filing years, a code was given to the current-year return and the prior-year return, using the current-year CCRS. County equivalent codes are assigned to the District of Columbia, the Virgin Islands, Puerto Rico, APO/FPO (military), and "other foreign."

▶ Identifying Migrants

Once the geographic codes are in place, Census determines who in the file has or has not migrated. The coded returns for 2 consecutive years are then compared to one another for two criteria: (1) the street address and (2) the mailing address State plus ZIP code. If the two are identical, the return is labeled a "nonmigrant." If any of the above information changed from the first prior year of study to the current year, the return is considered a mover. However, the return is only a "migrant" if the taxpayer's geographic code changed. If a taxpayer's address codes change from one year to the next, that taxpayer is an "in-migrant" for the address on the return filed in the second year, and an "out-migrant" for the address on the return filed the first year. If a taxpayer changed streets but stayed in the same county, that taxpayer would not be a migrant for purposes of this dataset.

As previously mentioned, the filer's return address determines the migration status of the record. There are instances, however, where the taxpayer may not have changed residences but the return address suggests a move. This may happen if: (1) the filing address is that of a financial institution or tax preparer, and not the actual taxpayer; (2) the taxpayer is a college student living away from home who filed with a home address one year and the college address another; (3) the taxpayer puts his or her place of business as the return address; (4) the taxpayer maintains dual residences, primarily residing in one county but having the tax return sent to the other; and (5) the taxpayer uses a post office box for mailing purposes.

▶ Tax Year versus Migration Year

This section distinguishes among what is meant by tax year, filing or calendar year, and migration year. When dealing with income taxes, the year in which a return is filed is the "filing" or calendar year and almost always follows the actual "tax year." For this reason, clarification of what exactly is meant by the year of migration is necessary. The residence of a taxpayer, for purposes of the Migration data files, is noted at the time the individual income tax return is filed. Because most tax returns are filed the spring after the tax year

has ended, the migration (filing) year coincides with the previous year's tax data. For example, the 2003 migration data cover the place of residence for individuals who were filing their 2002 Forms 1040 in Calendar Year 2003. Furthermore, since the migration data show movement from year to year, the files are expressed in 2-year increments, such as the 2002-2003 migration data. Thus, the file would show actual changes in residence from Calendar Year 2002 to Calendar Year 2003.

▶ IRS Preparation and Marketing of Migration Products

After Census geocoding and error checking, the Census Bureau maintains a file to supplement its internal population studies.[3] A copy is then delivered to the Statistics of Income (SOI) Division of the Internal Revenue Service. A statistician at the SOI Division checks the data for outliers, adds column headings and labels, and parses the data into Excel spreadsheets. Once SOI is satisfied with the dataset, it authorizes Census to release the file to State demographers. For each State, there is an inflow and an outflow spreadsheet, which shows the following information about the returns in each county: the number of migrant returns (used to estimate households); the number of exemptions attached to these returns (used to estimate individuals); the aggregate adjusted gross income of the migrating returns; and the median adjusted gross income of these returns. There is also a line item for nonmigrants with their relative incomes. An example of a page of the Minnesota inflow file for 2002-2003 follows (Figure A). This example shows the summary information for returns moving into Minnesota between 2002 and 2003, as well as detailed information

Figure A -- Inflow File for Minnesota (MN), 2002-2003

From St Abbr	From County Name	Number Of Returns	Number Of Exemptions	Aggregate Adjusted Gross Income (thousand dollars)	Median Adjusted Gross Income (whole dollars)
MN	Total Mig - US & For	146,999	257,176	5,894,696	25,079
MN	Total Mig - US	144,355	253,910	5,858,968	25,484
MN	Total Mig - US Same St	103,195	179,330	4,075,991	26,690
MN	Total Mig - US Diff St	41,160	74,580	1,782,977	22,294
MN	Total Mig - Foreign	2,644	3,266	35,728	4,877
MN	Aitkin County Tot Mig-US & For	454	875	18,991	28,102
MN	Aitkin County Tot Mig-US	454	875	18,991	28,102
MN	Aitkin County Tot Mig-Same St	393	767	16,643	28,599
MN	Aitkin County Tot Mig-Diff St	61	108	2,348	24,999
MN	Aitkin County Non-Migrants	5,175	11,257	200,253	25,733
MN	Hennepin County	58	105	2,833	38,332
MN	Anoka County	54	116	2,309	36,666
MN	Crow Wing County	47	91	1,627	18,999
MN	Ramsey County	29	52	1,640	45,832
MN	Itasca County	21	30	559	18,124
MN	Mille Lacs County	19	38	932	26,249
MN	Dakota County	18	32	964	37,499
MN	St Louis County	16	35	795	39,999
MN	Washington County	13	21	760	54,999
MN	Cass County	12	23	290	19,999
MN	Scott County	10	16	410	32,499
MN	Wright County	10	23	550	39,999
SS	Other Flows - Same State	86	185	2,974	24,999
DS	Other Flows - Diff State	61	108	2,348	24,999

for the first county of destination, Aitkin County, MN. For more information on interpreting this file, see IRS documentation. [4]

Once the files are prepared, they are announced for sale via the SOI Web site (www.irs.gov/taxstats/index.html), as well as in various SOI publications. The migration data are free to Federal, State, and local government agencies and are among the most popular products distributed through the SOI Division's Statistical Information Services (SIS) Office. In 2004, well over 200 migration data sets were distributed to customers in government, business, and academia. Information on pricing can be found on the Web site (www.irs.gov/taxstats/indtaxstats/article/0,,id=96816,00.html); in the Products and Services Section of each *Statistics of Income Bulletin*, Publication 1136; or by contacting the SIS office at (202) 874-0410.

▶ Strengths and Limitations of the Dataset

The county-to-county migration data may be the largest dataset that tracks movement of both households and people from county to county, including family incomes. Because these data are obtained from income tax records, they are inclusive and reliable. However, the source and design of this dataset have some limitations. As mentioned previously, those who are not required to file United States Federal income tax returns are not included in this file. Because of this, the dataset underrepresents the poor. Also not included is the small percentage of tax returns filed after late September of the filing year. Because the IRS granted most taxpayers who file this late an extension, and because most taxpayers who request an extension are more likely to file high-income tax returns, the migration data set can underrepresent the very wealthy.

The matching process also causes some returns to be missed. When the current-year tax return is compared to the prior-year tax return, only the Social Security number of the primary taxpayer is considered. If a secondary filer exists (as in the case of a married couple filing jointly), that Social Security number is not recorded or compared. If, for example, a husband and wife file a joint return in the prior year but file separately in the current year, only the husband's current year will have a match with the prior year. The spouse's current-year return becomes a nonmatch and will not be included in the data. This problem not only occurs when couples decide to switch filing status from year to year, but also when marriage or divorce changes an individual from being a primary taxpayer (included in the file) to a secondary taxpayer (not included in the file).

In addition to the dataset not including the entire individual filing population, it also underrepresents the elderly, another large segment of the population which may not be required to file individual tax returns.

▶ Uses of the County-to-County Migration Data

Statistics of Income tax data are mainly used within the Government by the Treasury Department's Office of Tax Analysis (OTA) and by the Congressional Joint Committee on Taxation. Both use the data in tax policy research and in revenue estimating.[5] The county-to-county migration data, however, are created for users outside the IRS or Treasury Department.

The Census Bureau uses these files to back up its demographic data between Decennial Censuses. Most of the individuals ordering these data are from academia, the media, and the private sector. Academic papers using the data show trends and shifts in demographics. Newspapers often highlight trends showing the fastest growing counties, where the wealthy are moving, and what parts of the country are losing population. Private firms include researchers hired by corporations, developers following movement of housing consumption, and technology companies estimating future demand, to name just a few. The county-to-county migration data are one of the most frequently requested products disseminated by the SOI Division. In Calendar Year 2004, the Statistical Information Services Office of the Division answered 367 requests about its migration data.

▶ Current Migration Trends

The wealth of useful data present in the county-to-county migration files can be illustrated by examining some current demographic trends shown in the data.

This section looks at three regional trends, as well as how customers used SOI data in their work.

► Loudoun County, Virginia

A look at inflow and outflow files for the State of Virginia shows that the fastest growing county in the Washington, DC metropolitan area is Loudoun County, Virginia. Loudoun County is situated just to the west of what used to be considered the outer limits of the Washington, DC suburbs as recently as 15 years ago. As the greater DC area continues to grow as a result of a long period of economic growth and small unemployment rate, more and more households have been moving into the area.

Two enormous residential communities, Ashburn and South Riding, evolved in the 1990's and are continuing to grow and attract affluent professionals by the thousands each year.

Figure B. -- Loudoun County, Virginia

	Number of Returns	Number of Exemptions	Aggregate AGI (thousand dollars)	Median AGI (whole dollars)
Inflows	13,073	27,035	939,231	50,864
Outflows	7,391	14,632	492,439	44,932
Nonmigrants	68,231	166,364	5,987,797	65,184

A look at the 2002-2003 data in Figure B compares the individual income tax return data of those who came into the county and those who exited the county between these 2 years. The Number of Returns column shows that the number of households increased by 7.5 percent between 2002 and 2003. The rise in number of exemptions nearly mirrors this change. A comparison of Adjusted Gross Income (AGI) between in the inmigrants and outmigrants is equally striking. The median AGI column shows that the median adjusted gross income of the returns moving into Loudoun County is considerably higher than the median income of those who are leaving. Both are lower than the median income of the nonmigrants (those who resided in Loudoun County for both years), suggesting that perhaps the inmigrants are younger and less-established families than those who have resided there longer.

► Clark County, Nevada

Another notable county in the United States in terms of migration is Clark County, Nevada. Clark County is the home of the cities of Las Vegas, North Las Vegas, and Henderson, as well as the unincorporated towns of Paradise (including the Las Vegas strip, the University of Las Vegas, and McCarran International Airport), Sunrise Manor, Spring Valley, and Enterprise. An examination of Figure C shows that, while 28,962 returns left the county from 2002 to 2003, some 44,311 returns came in. Thus, the returns moving into the county outpaced the returns leaving the county by 53 percent in that year. While Clark County is considered an excellent place to retire, data from the Nevada State Demographer's office show that the percentage of Clark County residents age 65 and older has held steady at approximately 11 percent for the past several years.[6]

Figure C. -- Clark County, Nevada

	Number of Returns	Number of Exemptions	Aggregate AGI (thousand dollars)	Median AGI (whole dollars)
Inflows	44,311	83,219	1,916,647	22,547
Outflows	28,962	54,254	1,028,971	21,010
Nonmigrants	511,010	1,084,081	25,334,202	32,015

The IRS county-to-county migration files also show that, of the top ten counties of origin for those moving into Clark County, none of them originates from the State of Nevada. The top five counties of origin are: Los Angeles, San Diego, Orange, and San Bernadino (all southern California counties), and Maricopa County, Arizona. Further study of the Nevada State Demographer's published data show that Clark County is projected to double in size between the years 2003 and 2024, accounting for 85 percent of the total expected growth in the State of Nevada for that time period.

▶ Riverside County, California

The U. S. county with the highest net gain of returns between Calendar Years 2002 and 2003 was Riverside County, California. Riverside County is situated just to the east of Los Angeles and Orange Counties, two of the most populated counties in Southern California. As shown below in Figure D, Riverside had a net gain of 20,404 returns during this time period. Where did these residents come from? According to the IRS data, 10,425 of the 50,843 returns coming in to Riverside County were former residents of Orange County. While having twice the population of Riverside County, Orange County is geographically small: only 789 square miles, compared to Riverside's 7,207 square mileage.

Figure D. -- Riverside County, California

	Number of Returns	Number of Exemptions	Aggregate AGI (thousand dollars)	Median AGI (whole dollars)
Inflows	50,843	114,863	2,282,503	30,189
Outflows	30,439	62,084	1,151,864	23,437
Nonmigrants	488,511	1,204,255	23,218,621	31,618

The second largest source of in-migrants to Riverside County was Los Angeles County, which lost 9,167 residents to this neighbor to the East. This loss may be a drop in the bucket for hugely populated Los Angeles, which has over 3 million residents, but illustrates a national trend: households are leaving the cities and close-in suburbs for more land and more affordable housing. In fact, Los Angeles had a significant net loss of households in the year examined, with 18,432 of its Year 2002 returns calling a different county home in 2003. The top five recipients of Los Angeles outflows were all neighboring Southern California counties.

▶ Summary

As this paper shows, the migration data contain a wealth of information that can be used to analyze and illustrate major demographic trends. The Census Bureau, in partnership with the IRS, creates a unique product rich in information yet simple enough to understand for all customers: from demographers, newspapers, and Government agencies to the public at large.

▶ Endnotes

1 U.S. Census Bureau, *Current Population Survey*, 2003 Annual Social and Economic Supplement.

2 *Statistics of Income--2002, Individual Income Tax Returns*, Publication 1304, Internal Revenue Service.

3 Long, John F, "Postcensal Population Estimates: States, Counties, and Places," presented at the Annual Meeting of the American Statistical Association, San Francisco, CA, August 1993.

4 Internal Revenue Service (1999), Area-to-Area Migration and County Income, internal documentation, Statistics of Income Division.

5 Kozielec, John (1996), "The Tax Return: A Unique Data Source for Tracking Migration," *Turning Administrative Systems Into Information Systems: 1995*, Publication 1299, Internal Revenue Service.

6 State of Nevada Demographer: "Nevada's Age, Sex, Race, and Hispanic Origin Estimates for 2003," http://www.nsbdc.org/demographer/pubs/.

▶ References

Sater, Douglas K. (1994), "Geographic Coding of Administrative Records--Current Research in ZIP/Sector-To-County Coding Process," working paper, United States Census Bureau.

United States Census Bureau (1996), Supplemental Documentation for External Data Products, internal documentation.