
Measuring the Impact of Interstate Migration on Federal Income Tax Receipts

Brian Francis and Emily Gross, Internal Revenue Service

Taxpayer migration data are some of the most frequently requested products disseminated by the Internal Revenue Service's Statistics of Income (SOI) Division. These data, including State-to-State and County-to-County migration, are produced from a collaborative initiative between IRS and the Census Bureau, using year-to-year changes in address information from individual income tax returns. This joint effort provides the Census Bureau with valuable income data and the IRS with a popular and useful data product. Migration data have been used by several newspapers to show population flows in their respective communities. Academic researchers also use these data to track migration patterns. This paper will begin with an overview of the IRS's Statistics of Income Program, which distributes the migration data, and the joint IRS-Census effort to produce them. This will be followed by descriptions of the migration data files and how they are created; the information they contain; and how the data are used. Next, there will be a brief examination of how the migration data can be used on national, State, and county levels as an application of using data products. Finally, the State-to-State migration data will be analyzed to determine the impact of interstate movement of taxpayers on Federal tax receipts for 1995-1999.

◆ The Statistics of Income Program

The Statistics of Income program began in 1916, when Congress passed a revenue act including 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. Specifically, the current Internal Revenue Code states that "the Secretary (of the Treasury) shall prepare and publish not less than annually statistics reasonably available with respect to the operations of the internal revenue laws..."[1].

Besides the annual SOI publications, based on individual and corporation 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. Also covered are studies on "international" tax returns, such as Controlled Foreign Corporations, foreign corporations with U.S. operations, foreign and U.S. possessions tax credits, Americans living abroad, and foreign recipients of U.S. income. Nearly all these statistics are based on samples of unaudited tax returns. In addition to the published reports, some SOI tabulations are also available on magnetic tape, diskettes, and CD-ROM. The SOI website (embedded in the IRS website) provides hundreds of downloadable spreadsheet files containing all of its current and recent publications [2].

Who uses SOI data? The main customer is the Treasury Department's Office of Tax Analysis (OTA), which decides on most of the program content, assisted by the Congressional Joint Committee on Taxation (JCT). Both use the data in tax policy research and in revenue estimating; however, they rely primarily on the data files that underlie the statistics, rather than on the published tabulations. As a result, the SOI files often contain more data than those published.

Other SOI customers are the Commerce Department, which uses tax return data in the National Income and Product Accounts, various other Federal agencies, State and local governments, universities, and the general public.

Most of the data are U.S. totals; not much is available by State and local area, except for the State summaries published in the *Bulletin* and the State and county data, which are the subject of this paper. At present, SOI does not have resources to produce geographic data.

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) pays IRS for annual data on every entity on the Individual

Master File; however, it must be able to justify the data items it receives as needed for its own statistical programs. SOI acts as the liaison for the IRS in its dealings with Census, and, as partial compensation, receives the migration data that Census produces based on tax returns.

◆ The Migration Data

How the File is Created

State-to-State and County-to-County migration flow data show migration patterns between U.S. States, based on year-to-year changes in addresses entered on income tax returns filed by individual taxpayers. The data are based on administrative records from the IRS's Individual Master File (IMF) system, which includes a record for every Form 1040, 1040A, 1040EZ, 1040PC, 1040SS, 1040PR, 1040C, and 1040NR tax return filed by citizens and resident aliens with the IRS [3].

Analysts at the Census Bureau match the Social Security number on tax records of individual income tax returns for adjacent years. Census checks the address, and, if it changes, the taxpayer becomes an "in-migrant" for the address on the return filed in the former year. Counts are made of the numbers of returns and "personal exemptions" claimed (exemptions can be used to approximate population counts), and aggregate adjusted gross income and median adjusted gross income [4]. (As a general rule, individual tax returns are filed during the spring following the end of the tax year. The 1998 tax returns are, for example, filed in the spring of 1999 and represent residence as of filing. Thus, when we refer to income or to the files themselves, we refer to tax year. When we refer to migration year, we refer to the year in which the return was filed. Thus, the match of Tax Years 1998 and 1999 produces 1999 and 2000 migration estimates.) Income tax is not one of the amounts included in the Census file because Census has not been able to justify receiving it. So, the migration data do not include tax.

When the file is completed, Census hand-delivers the migration files to SOI. A programmer at SOI then disclosure-proofs the data so that no taxpayer information can be calculated.

The migration data are marketed and sold by SOI. The data are usually distributed by e-mail or CD, and may be purchased by contacting SOI's Statistical Information Services office [5].

The migration data, like all tax return data, have shortcomings. Year-to-year matching is not fool-proof; for example, there can be problems associated with changes in name or marital status or the death of a taxpayer. In addition, because of the time of the year at which the files for Census are created, the Individual Master File provided to Census included only 95 percent to 98 percent of the returns filed that year. Moreover, not everyone has to file a tax return, especially if the individual (and spouse) is poor or elderly, with income below the filing requirements. [The IRS Area-to-Area Migration and County Income pamphlet describes the format and limitations of the data in more detail and includes pricing and ordering instructions. This pamphlet is available from SOI upon request.]

Notwithstanding the shortcomings of the return data, Census considers the IRS migration data to be the best source for estimating interstate migration; it also relies heavily on these data to estimate the population between censuses and to estimate per capita income.

Content

County-to-County and State-to-State Migration data are based on the Internal Revenue Service's Individual Master File records for all taxpayers. Up to 20 years of data are available (15 years for State-to-State Migration). Items provided include:

- Number of returns
- Number of exemptions
- Aggregate Adjusted Gross Income
- Median Adjusted Gross Income

A third product is the County Income Data, which do not show migration patterns, but do show expanded data for all counties in the United States. The County Income Data are also from the Individual Master File

record, and are available for the years 1989 to 2000. The items shown on this data set are:

- Number of returns
- Number of exemptions
- Aggregate wages and salaries
- Aggregate gross dividends
- Aggregate interest income
- Aggregate gross rents and royalties
- Aggregate adjusted gross income
- Median adjusted gross income

Some Migration Data Applications

American Demographics Magazine has included IRS as "one of the Best 100 Sources of Marketing Information," and notes that IRS migration data are the "finest information you'll find on annual migration rates for small areas, at least for the population under age 65." Apparently, many demographers and journalists agree. The County-to-County and State-to-State Migration Data are the most popular product sold at SOI's Statistical Information Services office. During the period April 2000 to April 2001, over 250 orders were filled for these files.

What are the data used for? In some cases, demographers are studying trends in migration. This may be used for academic research, for Government studies, or for private industry. Most frequently, the data are used for newspaper articles on moving trends, popular destinations for the wealthy, and areas of the country which show drastic changes in population over recent years.

Some of the trends are quite compelling. For instance, although many people realize that California is the most populous State in the country, many do not know that it has been experiencing a decline in population for more than 10 years straight. Which counties and States are welcoming these "out-migrants?" Three western counties head up the nation's county inflows:

Douglas County, Colorado (located halfway between Denver and Colorado Springs), Clark County, Nevada (Las Vegas, NV-AZ Metropolitan Statistical Area (MSA)), and Maricopa County, Arizona (in the Phoenix-Mesa MSA) [6].

In a paper presented at the Northeastern State Tax Officials Association Conference in 1995, County-to-County trends in the Northeastern States showed that although there were almost 509,000 people who moved into that region from the other 39 States outside of the Northeast, 726,000 people moved out of the region to any one of the other 39 States, a net loss of over 200,000 [7].

◆ **Impact of Migration on Federal Tax Receipts**

The next part of this paper conducts analysis using the Immigrants from all 50 States and the District of Columbia, for the years 1995 through 1999, and will attempt to show the effect that migration has on Federal tax receipts collected in each State. A regression model is tested which will allow us to determine the marginal impact of migrants. The effect of State unemployment will also be analyzed. Migration rates as opposed to the absolute number of migrants will be included in the model.

The Data

The variables in the data set are shown in Figure A. All data were obtained from SOI with the exception of unemployment rates, which were obtained from the Bureau of Labor Statistics. Descriptive statistics are presented in Figure B.

The model to be estimated is described below and was used to determine the effect of State-to-State migration on State Federal income tax receipts. This section of the paper will attempt to show the fiscal impact of migration by testing a multivariate regression equation using a panel data set comprised of the 50 States and the District of Columbia for the years 1995 through 1999. The basic model is as follows:

Figure A.—Names and Definitions of Regression Variables

Variable	Definition
State	Two-character postal abbreviation
Year	Tax year of returns used for migration
Tax	Federal income tax receipts
Migrants	Inflow returns
Exemptions	Exemptions claimed on migrants' returns
Migrants rate	Total migrants divided by total non-migrant returns
Exemption rate	Total exemptions divided by non-migrant exemptions
UE – Unemployment	Annual State unemployment rate
S – South	Binary variable equal to 1 if State is in the South, 0 otherwise.
N – Northeast	Binary variable equal to 1 if State is in the Northeast, 0 otherwise.
M – Midwest	Binary variable equal to 1 if State is in the Midwest, 0 otherwise.
W – West	Binary variable equal to 1 if State is in the West, 0 otherwise.

The Model

$$\text{Tax}_{ys} = \alpha_{ys} + \beta \text{Migrants}_{ys} + \beta \text{UE}_{ys} + \beta S + \beta N + \beta M + \varepsilon_{ys}$$

α_{ys} = an intercept

Tax = Federal income tax receipts (the dependent variable)

β = coefficients

Migrants = number of immigrants (the independent variable)

UE = State unemployment

ε = an error term that follows the first order autoregressive process

$$\varepsilon = \rho \varepsilon_{y-1} + \mu$$

y = 1995-1999

s = AL – WY (the 50 States)

Alternative Specification

$$\text{LTax}_{ys} = \alpha_{ys} + \beta \text{LMigrate}_{ys} + \beta \text{UE}_{ys} + \beta S + \beta N + \beta M + \varepsilon_{ys}$$

where LTax is the log of tax receipts and LMigrate is the log of the migration rate. The migration rate equals total migrants divided by total non-migrants for each State. The remaining terms are as defined above.

Regression Results

The first model is a regression of tax receipts on migrants, unemployment, and regional dummy variables. The R_2 for this model, 0.8648, shows that it fits the data well. All of the regressors are highly significant. Regression statistics are shown in Figure C.

The B-values are the coefficients which are the partial derivatives of the equation. That is, $\partial \text{Tax} / \partial \text{Migrants} = 36.5$. For each percentage change in the independent variable, the dependent variable changes correspondingly. Dividing by the quantities will give us the elasticity of tax revenue with respect to migrants. Using the means from the descriptive statistics, we compute the elasticity to be 5.715. Thus, each percentage increase in immigrants increases tax receipts by \$5,715. The positive coefficient for UE can be explained by the fact that unemployment compensation is taxable income.

In the next model, we convert the tax receipts and migration rates to logarithms so the coefficients are elasticities. Instead of the "level" of migrants, this model uses the rate of migration as a regressor. The negative coefficient on LMIGRATE shows that tax receipts fall as the rate of migration drops. The rate of unemploy-

Figure B.--Descriptive Statistics for Migration Data Panel, Filing Years 1995-1999

1995

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Migrants	51	59194.24	49873.47	9362.00	239487.00
Tax	51	1900648.80	2023480.62	179935.00	10648692.00
UE	51	5.2235294	1.2738270	2.6000000	8.9000000

1996

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Migrants	51	57687.39	49067.11	9251.00	228880.00
Tax	51	1875523.08	1988982.62	174841.00	10447492.00
UE	51	5.1549020	1.2546416	3.0000000	8.5000000

1997

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Migrants	51	59296.00	52457.64	9171.00	245522.00
Tax	51	1922666.78	2049540.55	179609.00	10851564.00
UE	51	4.7607843	1.2358120	2.5000000	7.9000000

1998

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Migrants	51	60103.96	53977.42	9275.00	246812.00
Tax	51	1882700.63	2019204.16	173309.00	10743774.00
UE	51	4.4000000	1.1963277	2.6000000	8.8000000

1999

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Migrants	51	61029.94	54013.39	8828.00	241352.00
Tax	51	1908465.67	2050905.41	176151.00	10956674.00
UE	51	4.1254902	1.0588377	2.5000000	6.6000000

ment moves inversely with this model since higher unemployment means less workers are contributing to the tax base. The negative relation of the migration rate implies that Federal tax receipts drop by roughly \$2,000 for each percentage point of the rate decrease. Figure D gives the results for this model.

Since the Durbin-Watson statistics in both models are very close to 2, auto-correlation is not a problem. The other common error problem in regression analysis is heteroskedasticity. To test for this problem, we saved the residual using it as the dependent variable and ran both models again. Using ordinary least squares, we

Figure C.—Regression Statistics for Model 1

Variable	DF	B Value	t Ratio	Approx Prob
Intercept	1	-2144569	-8.928	0.0001
MIGRANTS	1	36.482688	40.225	0.0001
UE	1	332452	8.292	0.0001
S	1	-446943	-3.722	0.0002
N	1	945703	6.608	0.0001
M	1	1198337	8.433	0.0001

Figure D.—Regression Statistics for Model 2

Variable	DF	B Value	t Ratio	Approx Prob
Intercept	1	7.723665	15.516	0.0001
LMIGRATE	1	-2.078620	-13.398	0.0001
UE	1	-0.064402	-1.496	0.1359
S	1	0.082803	0.634	0.5264
N	1	-0.763570	-4.375	0.0001
M	1	-0.757036	-4.182	0.0001

Figure E.—Ordinary Least Squares Test for Model 1

Variable	DF	Parameter Estimate	T for H0: Parameter=0	Prob > T
INTERCEP	1	76145	0.311	0.7558
MIGRANTS	1	-0.184456	-0.198	0.8434
UE	1	-22628	-0.542	0.5882
S	1	92739	0.748	0.4554
N	1	35815	0.245	0.8068
M	1	18285	0.123	0.9019

test for correlations with the independent variables. Results from the first model are shown in Figure E. The ordinary least squares results from the second model are shown in Figure F. None of the regressors in either model is significant; thus, we find no evidence of heteroskedasticity.

◆ Summary

The migration data are a very rich resource for users of SOI data. These include academic researchers, newspapers, other Government agencies, State and local governments, and others. By carefully processing

Figure F.—Ordinary Least Squares Test for Model 2

Variable	DF	Parameter Estimate	T for H0: Parameter=0	Prob > T
INTERCEP	1	0.008615	0.017	0.9862
MIGRATE	1	-0.000375	-0.002	0.9981
UE	1	-0.000990	-0.023	0.9816
S	1	-0.007959	-0.061	0.9513
N	1	-0.003421	-0.020	0.9843
M	1	-0.008132	-0.045	0.9640

IRS master files, the Census Bureau is able to produce a product for the general public. In this paper, we have provided a detailed description of the migration data and analysis of the effects of migration on Federal tax receipts. Additional research using the county-to-county migration files remains to be pursued.

◆ **Notes**

- [1] *U.S. Internal Revenue Code of 1986* (as amended by public laws subsequent to Public Law 99-514, October 22, 1986), Section 6108 (a), "Statistical Publications and Studies, Publication or Other Disclosure of Statistics of Income."
- [2] The web address for SOI is http://www.irs.gov/prod/tax_stats/index.html.
- [3] Forms 1040, 1040A, 1040EZ, and 1040PC are all variations of the standard form filed by individual United States citizens. Form 1040PR is for residents living in Puerto Rico; Form 1040NR is the individual form for nonresident aliens.
- [4] Internal documentation, Census Bureau, 2000.
- [5] For more information on SOI migration data, contact:

Statistical Information Services
 Statistics of Income Division
 P.O. Box 2608
 Washington, DC 20013-2608

(202) 874-0410
 sis@irs.gov

- [6] For more information on county migration trends in the Western U.S., see Pearson, Lucinda S.; Sater, Douglas K.; and Perry, Marc J. (2000), *Go East and North, Young Man: Destinations of Domestic Out-Migrants from California, 1989-1999*.
- [7] Koziolec, John, "The Tax Return: A Unique Data Source for Tracking Migration," presented at the Northeastern State Tax Officials Association Conference, Burlington, VT, October, 1995, and published in *Turning Administrative Systems Into Information Systems: 1995*, Publication 1299, Internal Revenue Service.

◆ **References**

- Internal Revenue Service (1999), *Area-to-Area Migration and County Income*, Internal Documentation, Statistics of Income Division.
- Perry, Marc J. (1998), "What's Going on in the Domestic Migration Pipeline? Recent Domestic Migration in Three Fast-Growing Counties," working paper, Southern Demographic Association.
- Maddala, G.S. (1988), *Introduction to Econometrics*, Macmillan Publishing Company, NY.
- Fomby, Hill, and Johnson, (1984), *Advanced Econometric Methods*, Springer-Verlag, NY. ■