ELECTRONIC DISSEMINATION OF INTERNAL REVENUE SERVICE LOCALITY DATA

Emily Gross and Beth Kilss,
Internal Revenue Service
Presented at the 2002 American Statistical Association

The Statistics of Income (SOI) Division of the Internal Revenue Service (IRS) offers many income tax data products in electronic format. While some of these products are electronic versions of SOI publications, others cannot be found in any published materials. By far, the most frequently requested of these types of data are individual taxpayer data by geographical area. The tables offered through SOI include: individual income and tax data by State and size of adjusted gross income; county income data; county-to-county migration data; State-to-State migration data; and ZIP code area data. Some of these tables are available on SOI's website as downloadable spreadsheet files, while others are disseminated by CD-ROM or e-mail.

This paper will begin with an overview of the IRS's Statistics of Income Program, which distributes the locality data. Next, there will be a description of the types of locality income data the IRS offers, the electronic formats available, the years covered by the data, and the sources and limitations of the data. Following that, practical applications of the data will be shown, as well as some interesting ongoing studies by some of our customers. Finally, some innovations for the future will be highlighted as the SOI Division plans to make its statistics even more readily available to the thousands of citizens who download its files each year.

Background on the Statistics of Income Program

The IRS's Statistics of Income (SOI) Program began over 86 years ago, following the passage of the Revenue Act of 1916, which required the annual publication of statistics. Although there have been many revisions to the tax law since that time, the original mandate of that Act continues today. Thus, throughout its long history, SOI has been preparing and publishing financial statistics from various types of tax and information returns filed [1].

SOI conducts the Internal Revenue Service's studies on individuals, corporations, partnerships, sole proprietorships, estates, nonprofit organizations, and trusts, as well as specialized studies covering both inbound and outbound international activities. Its mission is to collect and process data so that they become meaningful information and to disseminate this information to its customers and users.

Customers of SOI Data

SOI provides its data to many customers, but its two primary ones are the Office of Tax Analysis (OTA) in the Office of the Secretary of the Treasury, and OTA's legislative counterpart, the Congressional Joint Committee on Taxation (JCT). In the Department of Commerce, the Bureau of Economic Analysis (BEA) and the Bureau of the Census are significant users; the former uses the SOI sample data (content rich but lower volume) for estimating components in the national income and product accounts, and the latter's needs are primarily for master file ("population") data.

SOI data users outside of the Government sector include academia, the media, demographers, and public libraries. In addition, other areas of the IRS use SOI data for their internal operations.

Some SOI Products and Services

As a byproduct of producing data for its primary customers, Statistics of Income information is also made publicly available through both printed publications and electronic media. The *Statistics of Income (SOI) Bulletin* is published quarterly, with each issue containing four to eight articles and data releases of recently completed studies, as well as historical tables covering a variety of subject matter, from Treasury Department tax collections to taxpayer assistance and tax return projections [2]. SOI produces separate annual "complete reports" on individual and corporation income tax returns, which contain more comprehensive data than those published earlier in the *Bulletin* [3,4]. The *Corporation Source Book* is also published annually, presenting detailed income statement, balance sheet, and tax data by industry and asset size [5]. Another annual SOI product is the *IRS Data Book*, which contains statistical tables and organizational information on a fiscal year basis [6]. A series of reports, including research articles documenting technological and methodological changes in SOI programs and other related statistical uses of administrative records, is also published by SOI annually or biannually [7]. And last, but not least, periodically, SOI produces special compendiums of research and analysis, covering topics such as nonprofit organizations, estate taxation and personal wealth, and international business activities [8,9,10].

Public awareness of SOI products and services has grown rapidly over the past several years due to the IRS World Wide Web site, which provides users an easy option for accessing SOI data. The SOI web site, accessible via the "Tax Stats" option on the IRS home page, provides over 1,400 downloadable spreadsheet files containing all of SOI's current and recent publications. At present, close to 200,000 files are downloaded monthly from Tax Stats [11].

Other electronic media products are available from SOI on magnetic tape, CD-ROM, diskette, and files sent via e-mail. These products include the Individual Public-Use Microdata File (for which taxpayer identifiers have been removed); Exempt Organizations and Private Foundations Microdata Files (whose returns are open to the public); the *Corporation Source Book*; individual income tax return data shown by State, county, or ZIP code; and individual migration data shown on either a State or county basis. SOI's Statistical Information Services Office has staff available to facilitate the dissemination of SOI data, as well as to answer questions about the data [12].

IRS Locality Data

Of all the data products available from SOI, individual taxpayer data by geographic area are among the most frequently requested. Some of the State data are available online; most of the other geographic data are described online but must be obtained directly from SOI. The following summarizes the types of electronic locality data available and the particulars about each file:

Individual Income and Tax Data, by State and Size of Adjusted Gross Income, Tax Years 1997-2000

These tables present aggregates of all returns filed and processed through IRS's Individual Master File (IMF) system for Calendar Years 1997-2000. Classification by State was usually based on the taxpayer's home address. However, some taxpayers may have used the address of a tax lawyer or accountant or the address of a place of business. These tables are available through the IRS web site, www.irs.gov, by selecting "Tax Stats" in the upper left-hand corner, then selecting "Statistics by Topic." Under that heading, select "Individual Tax Statistics," then select "State Income" under the heading "Data by Geographic Areas" [13]. The tables shown are an expanded version of Table 2, which appears in the *SOI Bulletin* every spring, and include additional individual income and tax items, as well as seven adjusted gross income (AGI) size classes. All are viewable at the Internet site.

The AGI classes include the following: Under $20,000; $20,000 under $30,000; $30,000 under $50,000; $50,000 under $75,000; $75,000 under $100,000; $100,000 under $200,000; and $200,000 or more. There is also an "all returns" column with total numbers for each tax category.

The following are some of the 57 income and tax categories reported on our individual income tax data by State tables: number of returns, number of exemptions, adjusted gross income (total and components), total itemized deductions (total and components), total credits (total and components, including the earned income credit and the child care credit), and tax liability, taxes due, and overpayments.

County Income Data

These data are based on the population of returns from the IRS IMF system. They are presented in a table, by total "money income" (a Census Bureau income concept), adjusted gross income, wages and salaries, dividends, interest income, gross rents and royalties, number of returns (which can be used to approximate the number of households), and number of personal exemptions (which can be used to approximate the population). The data are presented by county (including State totals) and are available for 1989-2000 for the entire United States. A sample may be viewed at www.irs.gov/taxstats. As in the previous example, go to "Data by Geographic Area" and click on "County Income."

ZIP Code Area Data

Statistics are available for Income Years 1997 and 1998 on CD-ROM. The file is organized into 18 columns. Selected income and tax items are tabulated by State, ZIP code, and size of AGI. The AGI size classes included in this file are: Under $10,000 (which includes deficit returns); $10,000 under $25,000; $25,000 under $50,000; and $50,000 or more. Data items include the number of individual income tax returns; the number of personal exemptions; adjusted gross income; salaries and wages; taxable interest; earned income credit; total tax; number of returns with Schedules C and F and number of Schedules C and F; and number of returns with Schedule A and amount of Schedule A deductions, by State and 5-digit ZIP code. Examples of the 1997 and 1998 data are viewable from the Tax Stats web site, under "Data by Geographic Areas" as noted earlier, along with ordering information. Data for 1991 are also available at no charge at the same location on the Internet site.

State-to-State Migration Data

These data show migration patterns by State for the entire United States, including inflows and outflows,

and are based on year-to-year changes in the addresses shown on the population of returns from the IRS IMF system. The data include the number of returns, the number of personal exemptions, total adjusted gross income (starting with 1995-1996), median adjusted gross income (starting with 1995-1996), total money income (for years 1992-1993 through 1995-1996), and median total money income (for years 1992-1993 through 1995-1996). The data for each State are represented by two files: an inflow file and an outflow file. The inflow file shows all the above data for the total number of returns "entering" the State (i.e., the taxpayer resided in the State in the second year, but not in the previous year), as well as a migration breakdown of each tax item for each State of origin, in descending order. The outflow file shows the exact opposite for each State, depicting the number of returns leaving a State, as well as the destination States. The State-to-State migration data are available for 1989-2001. Click on "State to State Migration" at the same site path detailed above to view a sample of the data and for information on how to obtain them.

County-to-County Migration Data

Also based on year-to-year changes in the addresses shown on the population of returns from the IRS IMF system, these data show migration patterns by county for the entire United States, including inflows and outflows. The data include the number of returns, the number of personal exemptions, total adjusted gross income (starting with 1995-1996), median adjusted gross income (starting with 1995-1996), total money income (for years 1992-1993 through 1995-1996), and median total money income (for years 1992-1993 through 1995-1996).

The county-to-county migration product is also arranged by State, with each State represented by an inflow and outflow file. The county migration data show each county alphabetically, followed by the counties of origin (inflow file) or the counties of destination (outflow file). This product is available for 1984-2001. A sample is viewable by clicking on "County to County Migration" at the same location as noted above, which also has ordering information.

Practical Applications of IRS Locality Data

The locality or geographic data that SOI makes available online or through its SIS office has many practical applications. In some cases, demographers are studying trends in migration. These data 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 Customers and How They Use Our Locality Data

IRS locality data customers are using the various geographic products in a variety of ways. Some examples include: the media reporting on population and moving trends; businesses conducting research on income trends in localities to conduct their own business; universities studying population or income changes; and State and local governments examining how their tax bases are changing.

A majority of the users of these data are "regulars"--they usually order the most current data every year, either to be included in longitudinal studies or to offer to their clients. Below are examples of how three of our regular customers use the locality data:

- *The Hamilton County Regional Planning Commission, Hamilton County, OH*

  Hamilton County, OH, is the central county of the Cincinnati Consolidated Metropolitan Statistical Area (CMSA, a Census Bureau concept). For over 15 years, the commission has been using IRS county-to-county migration data to report on population and income trends in their area. The commission's report, "*Spreading Out: The March to the Suburbs*," shows population changes and migration to and from Hamilton County, OH, during the period of 1984-1997 [14]. (See Chart 1.)

  The commission was charged with this assignment because Cincinnati experienced a significant decline in population even though the regional economy was strong. The goal of the report was "to create a long-range, comprehensive, countywide plan to achieve optimal sustainable development and redevelopment in Hamilton County, while identifying the county's growth opportunities in the context of the region."

  This report examines growth trends in Hamilton County, population movement between the county and the other Cincinnati CMSA counties, interstate migration patterns, average household size of migrants and non-migrants, per capita income of migrants, total net income flows due to migration, and migration patterns from one metropolitan area to another. All of this can be gleaned from the IRS county-to-county migration data.

The detailed analysis had four conclusions. First, Hamilton County is losing population because people are physically leaving the county, not because the death rate exceeds the birth rate. Second, residents who leave the county are generally going to neighboring counties within the Cincinnati CMSA. Third, the typical non-migrant of Hamilton County generally has a higher income and larger household. Finally, the movement of people away from Cincinnati is "a trend, and not a fluke." (See Map 1.)

- *SMR Research Corporation, Hackettstown, NJ*

SMR is a business research firm that studies consumer financial services markets and the companies that participate in them. They are the nation's largest publisher of strategic research studies on consumer financial topics and also provide consulting, electronic data, and statistical services.

SMR provides corporations with data and analysis on U.S. consumer loan markets, financial services companies, and consumer wealth and credit risk by geographic area [15]. SMR uses SOI data for the nation with county and ZIP code to estimate the personal savings rate (PSR), a calculation commonly used in Economics. According to Stuart Feldstein, the president of SMR Research Corporation, "The percentage of people who report taxable interest income is a fine surrogate for the percentage of people who happen to have money in the bank." Mr. Feldstein continues to say that, tracked over time, this figure is as useful as the Commerce Department's PSR, which is used by virtually all economists.

For SMR, the county-level income data serve two purposes. They show mean incomes by county per capita (exemption) and by filer. This shows, at a local level, where incomes are rising, where they are falling, and by what amounts. SMR also uses the county fields of interest and dividend income, which Mr. Feldstein says are very valuable to banks and investment companies.

- *Office of the City Controller, Philadelphia, PA*

The City of Philadelphia, Office of the City Controller, is the sole auditing agency of Philadelphia City government. This role was expanded in 1965, also making the City Controller the auditor for the School District of Philadelphia. The City Controller is elected to a 4-year term midway between the elections of the Mayor and City Council to encourage a measure of independence from the officials whose expenditures the Controller's Office must audit.

One of the many services the City Controller's office provides is the Preferred Place Index (PPI). The PPI includes data on economic, social, and government trends, such as city wage and tax revenues, construction permits, migration trends, air quality, crime rates, and infant mortality. This index provides a basis for comparing a broad range of statistics on Philadelphia from year to year. It is generally considered a way to track the city government's performance.

One of the measures in the PPI is net migration within the Philadelphia Metropolitan area. Recognizing the important role that immigrants play in the growth of many American cities, a measure of the net number of immigrants in Philadelphia was included. Using IRS county migration data, the Controller's office identified those migrants with foreign addresses as either their origin or destination addresses and calculated the net number of immigrants. This factor was a weighted one because it is a subset of the total overall migration trend numbers and because of the significant potential for immigrants to fail to be counted by IRS data due to low income levels and other limitations [16].

The three examples above show how our locality data can track trends in population, income, financial markets, and even a city's likelihood to attract and retain residents. The media also use IRS locality data regularly when reporting moving trends and highlighting desirable communities.

Innovations for the Future

Future improvements of our locality data focus on making current data available sooner. As requests for the ZIP code data continue to rise, the SOI Division will focus on releasing the data more quickly. The 1999 and 2000 ZIP code files may be available by the end of 2002 and the beginning of 2003, respectively. The 2001-2002 County-to-County, State-to-State, and County Income Data files are expected to be available around the beginning of 2003.

The Division also plans to have more "user-friendly" locality data historical products. Customers who now

need to order separate files for longitudinal studies will find it easier to order one product, which contains several years of data, including the necessary documentation to interpret the changing file layouts (especially useful for the earlier years of the locality data).

The vast majority of SOI's customers receive their data via the "Tax Stats" web site. Plans continue to improve the search engine and make the web site even easier to use. The web will also provide a way to let more of the public know about our locality data.

Notes and References

[1] For further information on the Statistics of Income Program, see Petska, Tom, "'Statistical Information from Administrative Records in the Federal Tax System," *Special Studies in Federal Tax Statistics: 2000-2001*, Publication 1299, Internal Revenue Service. See also the SOI web site: www.irs.gov/taxstats

[2] See, for example, *Statistics of Income Bulletin*, Winter 2001-2002, Publication 1136, Internal Revenue Service.

[3] See, for example, *Statistics of Income--1999, Individual Income Tax Returns*, Publication 1304, Internal Revenue Service.

[4] See, for example, *Statistics of Income--1999, Corporation Income Tax Returns*, Publication 16, Internal Revenue Service.

[5] See, for example, *Source Book of Statistics of Income--1999, Corporation Income Tax Returns*, Publication 1053, Internal Revenue Service.

[6] See, for example, *Data Book 2001*, Publication 55B, Internal Revenue Service.

[7] See, for example, *Special Studies in Federal Tax Statistics: 2000-2001*, Publication 1299, Internal Revenue Service.

[8] See, for example, *Statistics of Income--Compendium of Federal Estate Tax and Personal Wealth Studies*, Publication 1773, Internal Revenue Service.

[9] See, for example, *Statistics of Income--Compendium of Studies of International Income and Taxes, 1984-1988*, Publication 1267, Internal Revenue Service.

[10] See, for example, *Statistics of Income--Compendium of Studies of Tax-Exempt Organizations, 1974-1987; 1986-1992;* and *1989-1998*, Publications 1416, Internal Revenue Service.

[11] The web address for SOI is www.irs.gov/taxstats. There is also a direct link to our web site from the Federal Government's "One-Stop Shopping for Federal Statistics": www.fedstats.gov .

[12] For more information on SOI products and services, including SOI locality data, contact:
Statistical Information Services
Statistics of Income Division
PO Box 2608
Washington, DC 20013-2608
(202) 874-0410
sis@irs.gov

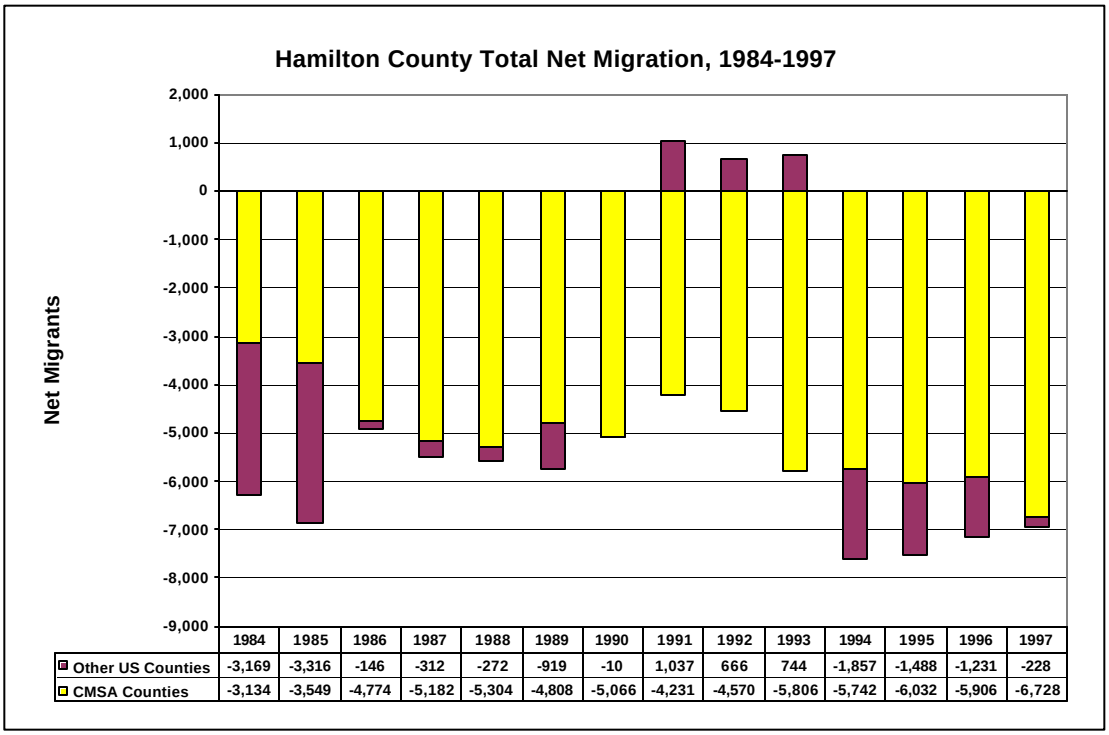
[13] At this location, the Tax Year 2000 expanded unpublished version of the individual income tax return data by State follows five tables for Fiscal Year 2001 from the *Internal Revenue Service Data Book*.

[14] Nolan, Christine (1999), "*Spreading Out: The March to the Suburbs*," Hamilton County Region Planning Commission, Cincinnati, OH.

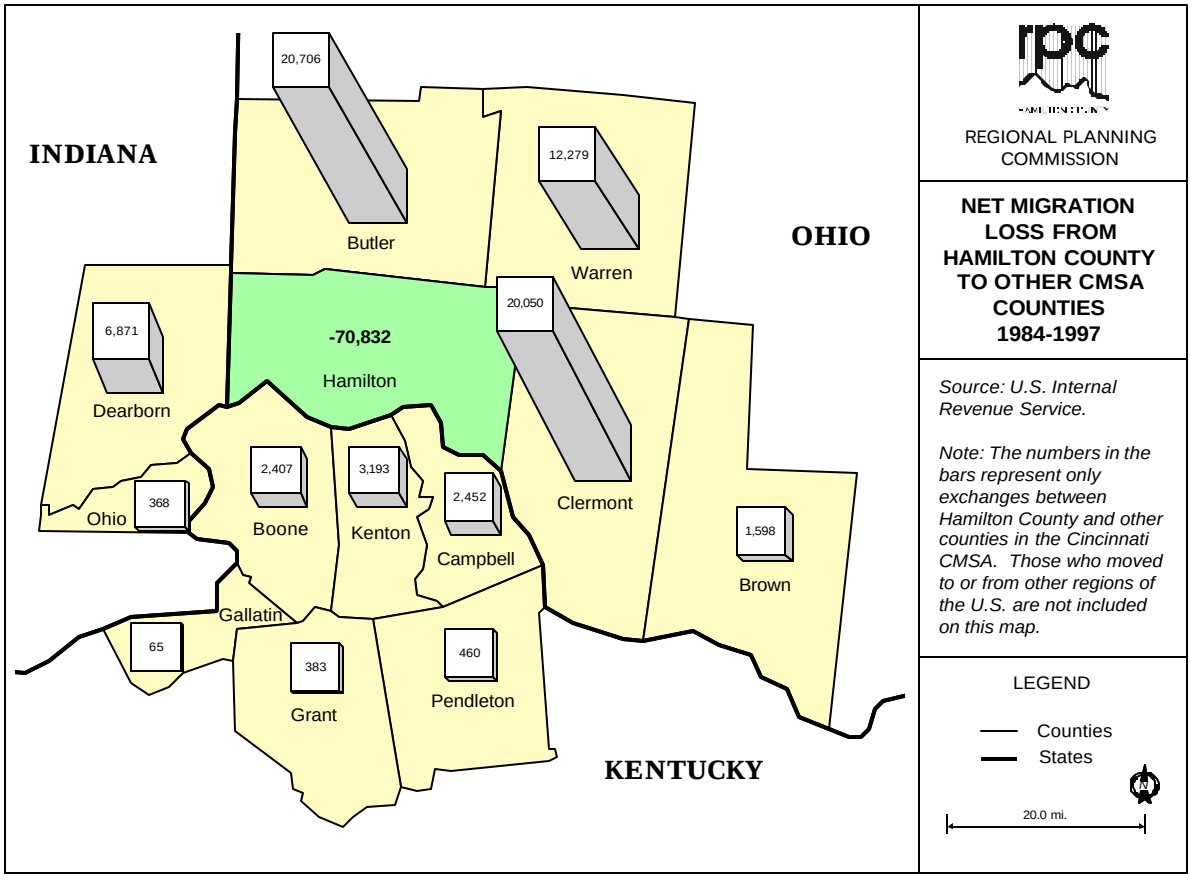
[15] For explanations of SMR Research Corporation's various products on U.S. consumer loans and other financial topics, see www.smrresearch.com.

[16] "The City Controllers Preferred Place Index," www.philadelphiacontroller.org.

Chart 1



	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Other US Counties	-3,169	-3,316	-146	-312	-272	-919	-10	1,037	666	744	-1,857	-1,488	-1,231	-228
CMSA Counties	-3,134	-3,549	-4,774	-5,182	-5,304	-4,808	-5,066	-4,231	-4,570	-5,806	-5,742	-6,032	-5,906	-6,728

Source: IRS Statistics of Income Division, 1999, prepared by Christine Nolan and the Hamilton County Regional Planning Commission.

Map 1