A Comparative Analysis of Income Statistics for the District of Columbia

ACS Income Estimates vs. DC Individual Income Tax Data

Jayron Lashgari

Introduction

The U.S. Census Bureau American Community Survey (ACS) is an important source of annual data on the economic, employment, housing, and demographic conditions of the nation and its subnational jurisdictions. The survey not only helps Americans know more about their communities, the information provided by the ACS helps local officials, community and business leaders understand important changes taking place in their communities.

Local jurisdictions also maintain a vast store of administrative data that captures some of the same socioeconomic measures included in the ACS. This raises the issue of how these measures from alternative data sources compare. In an attempt to address this issue, this analysis compares selected income related statistics for the District of Columbia presented by the ACS and District of Columbia individual income tax returns. The objectives of this analysis are to a) determine whether there are substantial differences between the comparable statistics from the two data sources; b) determine whether such differences are systematic; and c) attempt to reconcile such differences between comparable statistics from the two data sources. This analysis focuses on income measures in both the ACS and District of Columbia income tax returns database; it also compares the Gini coefficient measure in the ACS to one derived from income measures in the District of Columbia income tax returns database.

Data

American Community Survey Estimates

The ACS is an ongoing survey that provides vital information about our nation and its people. It collects data on households in a wide range of areas such as employment, educational attainment, and income and averages the data in the form of one-year, three-year and five-year estimates. This analysis used the one-year and five-year estimates for select income variables for the District of Columbia. In the case of the District of Columbia, the survey is designed to annually collect household data from about 4,000-5,000 city household respondents on their prevailing demographic and economic circumstances.

One year estimates are based on a sample size of about 4,000 to 5,000 household respondents, and the five-year estimates are based on a sample size of about 20,000 to 25,000 household respondents. The five-year estimates are the most reliable because of the larger survey sample size. However, it is based on not-so-current information. The one-year estimates, which have the smallest sample size, contain most recent information but are less reliable than the five-year estimate (because of the relatively small survey sample size).

ACS five-year income estimates are based on the most recent household survey responses for a given jurisdiction collected in a 60 month time period. ACS five-year income estimates group the household survey responses into the following respective five-year periods: 2006-2010, 2007-2011, 2008-2012, and 2009-2013. For example, all survey response data collected for years 2006 to 2010 is used to produce one average estimate for the time period 2006-2010. The self-reported survey responses are then used by the U.S. Census Bureau to help estimate the city aggregates for respective data variables.

The exact variable definitions used by the ACS are also important. For example, the ACS defines household income as the income of all individuals 15 years and older in a single residence

regardless of their relation to the homeowner. Household income comprises wages and salaries, military pay, commissions, tips, cash bonuses, social security payments, pensions, child support, public assistance, annuities, money derived from rental properties, interest and dividends. Earnings, as defined by the ACS, represent income received by individuals over 16 years old. Earnings, a narrower measure of income, are simply wages and salaries from employment and self-employment income.

ACS also produces one-year estimates of the Gini index of income inequality. The income Gini coefficient is a measure of the deviation of the distribution of income among households in the city from a perfectly equal distribution. A value of 0 indicates absolute equality (e.g. 50 percent of the households have 50 percent of the income), and a value of 100 indicates absolute inequality (e.g. 1 percent of the households have 100 percent of the income).

District of Columbia's Individual Income Tax Data

The District of Columbia's individual income tax data is collected and administered by the District of Columbia's Office of Tax and Revenue (OTR). The District of Columbia individual income tax data consists of the individual income tax return data for each resident tax filer in a given year. Each record for a given year contains pertinent tax and income information for each filer. The annual number of administrative tax records has grown from approximately 270,000 in 2001 to over 330,000 in 2013, reflecting population growth over the period. The analysis uses the Federal Adjusted Gross Income (FAGI) and the Wages, Salaries, and Tips variables from the individual income tax database as its primary measures of income variables. FAGI is comprised of wages, salaries, and tips, business income, income from rental property, and capital gains. The FAGI of all tax filers is deemed the comparable income measure for the ACS's "household income". Tax data "wages, salaries, and tips" is deemed the comparable income measure for the ACS's "earnings". The citywide FAGI data as stated on each filer's tax return was used to determine the Gini coefficients for the city for each year from 2006 to 2012.
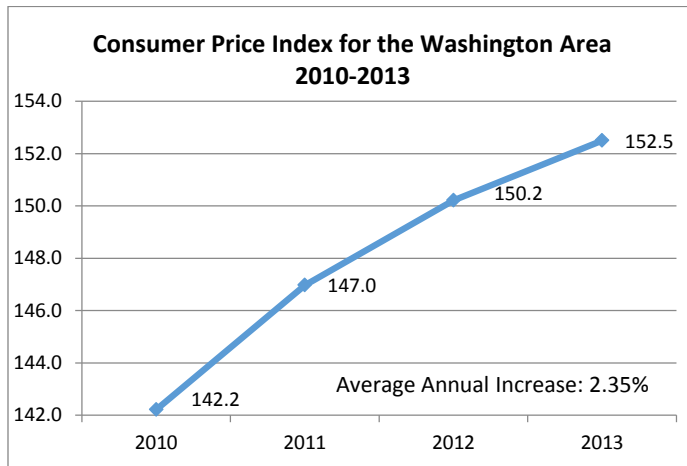
Comparing DC Income Tax Statistics to ACS Income Estimates

ACS five-year income estimates group the data (annual survey responses) into five-year periods: 2006-2010, 2007-2011, 2008-2012, and 2009-2013. In order to accurately compare the District's income tax data to the ACS data, the annual tax data was also grouped into the identical five-year periods. Each record in the tax data (just as each survey response in the ACS) was then adjusted for inflation to the same benchmark year as the ACS data. Once the tax data was organized in comparable five-year panels, the mean and median FAGI and wages, salaries, and tips were calculated to compare to the ACS estimates.
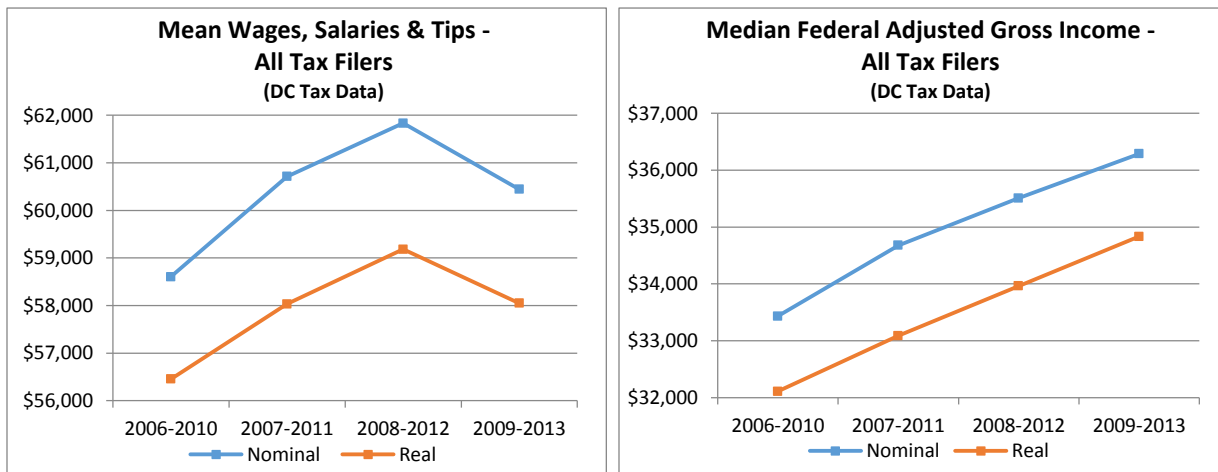
The ACS five-year income estimate for each five year period was adjusted for inflation. For example, all survey income responses collected by the survey between years 2006 and 2010 were put in 2010 real dollar terms, all survey income responses collected by the survey between years 2007 and 2011 were put in 2011 real dollar terms and so on. The same adjustments were made to the annual tax data. The U.S. Bureau of Labor Statistics states that for the Washington-Baltimore metropolitan area, the consumer price index increased from 142.2 in 2010 to 152.5 in 2013 as shown in Figure 1. The index value for the metro area increased by 3.3 percent in 2011, 2.2

percent in 2012 and 1.5 percent in 2013. Each tax record in any given year was adjusted for inflation by the appropriate deflator. As a consequence, the real values of the mean wages, salaries and tips for all tax filers are about 4 percent lower than the nominal values as shown in Figure 2. The real values of the median federal adjusted gross income for all tax filers are also about 4 percent lower than their respective nominal values.

Figure 1: Washington Area CPI: 2010 - 2013

Consumer Price Index for the Washington Area 2010-2013

Average Annual Increase: 2.35%

Year	CPI
2010	142.2
2011	147.0
2012	150.2
2013	152.5

Figure 2: Nominal and Real Income Levels for The District of Columbia

Mean Wages, Salaries & Tips - All Tax Filers (DC Tax Data)

Nominal, Real — for periods 2006-2010, 2007-2011, 2008-2012, 2009-2013

Median Federal Adjusted Gross Income - All Tax Filers (DC Tax Data)

Nominal, Real — for periods 2006-2010, 2007-2011, 2008-2012, 2009-2013

Income and Gini Measures: ACS versus Income Tax Data

Income Levels

For all time periods under investigation, median earnings for District workers in the ACS survey was at least $8,000 (20 percent) higher than the median wages and salaries in the DC tax returns database (Table 1). The responses to the ACS survey indicated that the citywide median earnings in the 2009-2013 time period was $45,231, while the actual median wages and salaries of all tax filers was $36,288 for the same period. In the cases of the ACS and DC tax data, the median was derived by taking the middle value of inflation adjusted income data for the respective 60-month time period. Turning attention to household income, a slightly broader definition of income, in Table 2 and Figure 3, the ACS data states that median household income for the city tended to be more than $16,000 (30 percent) higher than the median FAGI for all tax filers in the city. The responses to the ACS survey indicated that the citywide estimated median household income in the 2009-2013 time period was $65,830 compared to the $44,794 actual median income of all tax filers.

Table 1: Median Wages/Earnings (dollars)

	2006-2010	**2007-2011**	**2008-2012**	**2009-2013**
Median Earnings for Workers (ACS)	$41,171	$43,137	$44,423	$45,231
Median Wages, Salaries & Tips (DC tax data)	$33,432	$34,679	$35,510	$36,288
Amount Difference	$7,739	$8,458	$8,913	$8,943
% Difference	20.8%	21.7%	22.3%	21.9%

Table 2: Median Income (dollars)

	2006-2010	**2007-2011**	**2008-2012**	**2009-2013**
Median Household Income (ACS)	$58,526	$61,825	$64,267	$65,830
Median FAGI (DC tax data)	$42,017	$43,317	$44,124	$44,794
Amount Difference	$16,509	$18,508	$20,143	$21,036
% Difference	32.8%	35.2%	37.2%	38.0%

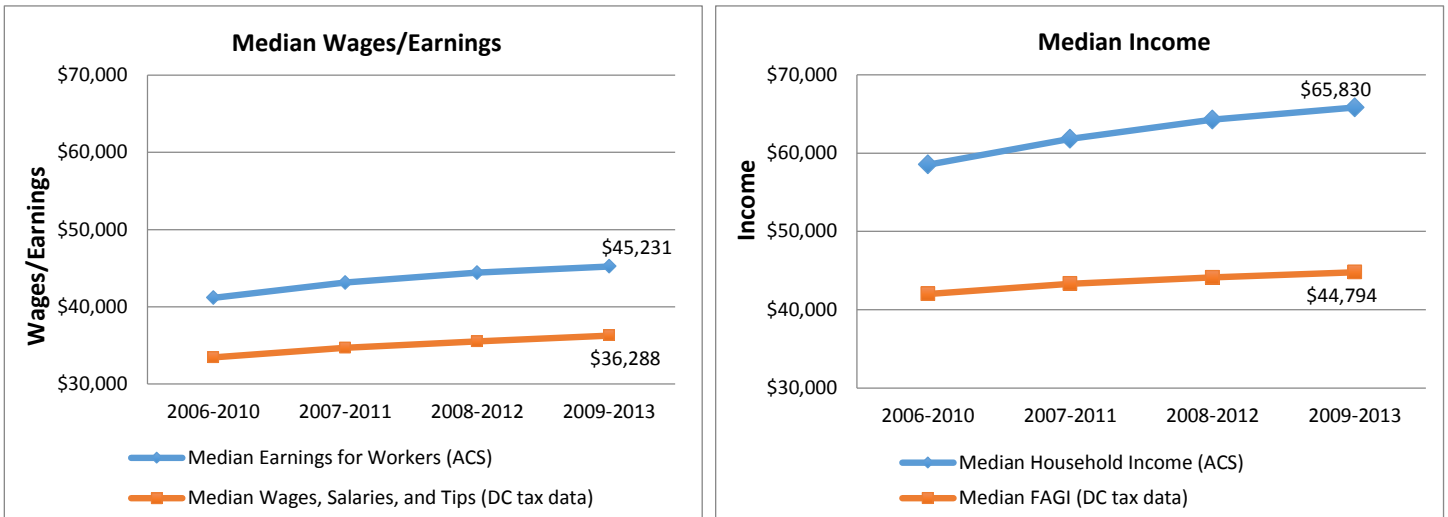**Figure 3: Median Wages/ Earnings: ACS versus Tax Data**

Figure 3 shows persistent and fairly constant differences between the median measure of earnings in the ACS and the median measure of wages in the tax data. But upon closer inspection, Figure 3 shows a mildly increasing difference between the median measures of income in the ACS versus the administrative data.

Figure 4 and Table 3 show a more noticeable growing difference between measures of mean income in the two data sources. The ACS mean income in the 2006 – 2010 time period was $8,189 (9.3 percent) higher than the comparable statistic obtained from local income tax data. By the 2009 – 2013 time period the ACS mean income was $17,589 (19.2 percent) higher than the mean income measure obtained from the tax data. Over the period, the mean FAGI decreased by an estimated annual rate of 0.04 percent compared to a 3.27 percent estimated annual growth rate for mean household income in the ACS.

In addition to the mean incomes, Figure 4 displays the 90 percent confidence interval for the ACS means. The ACS sample estimates and their statistical standard errors (the basis of the upper and lower bounds of the confidence interval) allow for the construction of confidence intervals. The interval represents the degree of certainty about the point estimate. The interval can be interpreted as providing 90 percent certainty that the interval defined by the upper and lower bounds contains the true value of the statistic under investigation. In practice this means if the ACS replicated their processes and methodology used to obtain a given statistic 100 times, the resultant statistics are expected to be within the said confidence interval 90 percent of the time. Figure 4 suggests that the process and method by which OTR obtains and processes its administrative tax data is considerably different from that of the ACS.

Figure 4: Mean Income: ACS versus Tax Data
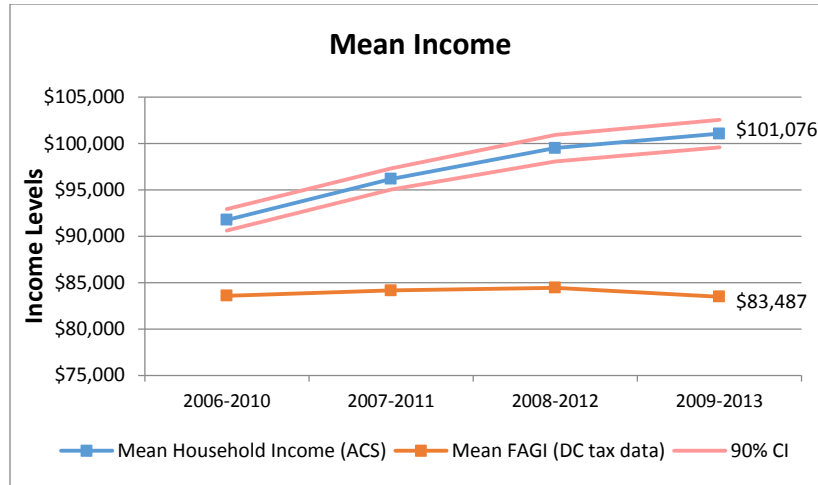
Table 3: Mean Income (dollars)

	2006-2010	2007-2011	2008-2012	2009-2013
Mean Household Income (ACS)	$91,778	$96,183	$99,511	$101,076
Mean FAGI (DC tax data)	$83,589	$84,161	$84,448	$83,487
Amount Difference	$8,189	$12,022	$15,063	$17,589
% Difference	9.3%	13.3%	16.4%	19.2%

Table 4: Growth in Mean Income

	2006-2010	2009-2013	Estimated Annual Average Growth
Mean Household Income (ACS)	$91,778	$101,076	3.27%
Mean FAGI (DC tax data)	$83,589	$83,487	-0.04%

Income Inequality

Gini coefficients reported by the two sources share a similar profile but differ in three important ways (Figure 5). First, the Gini coefficients produced by the ACS are 12 to 18 percent lower than that produced using income tax data. This could significantly affect one's view of inequality in the city. Second, the ACS coefficients stay in the very tight range of 0.53-0.54 whereas the tax data coefficients range from 0.61 to 0.66. These ACS coefficients suggest that the level of income inequality has been relatively unchanging whereas the tax data coefficients show an appreciable decline as the great recession began to take its toll on the city's economy. Third, the tax data show income inequality began to increase in 2012 while the 2012 ACS coefficient appears practically indistinguishable from the coefficients in 2009 to 2011.

Figure 5: Gini Coefficients for the District of Columbia: ACS versus Tax Data
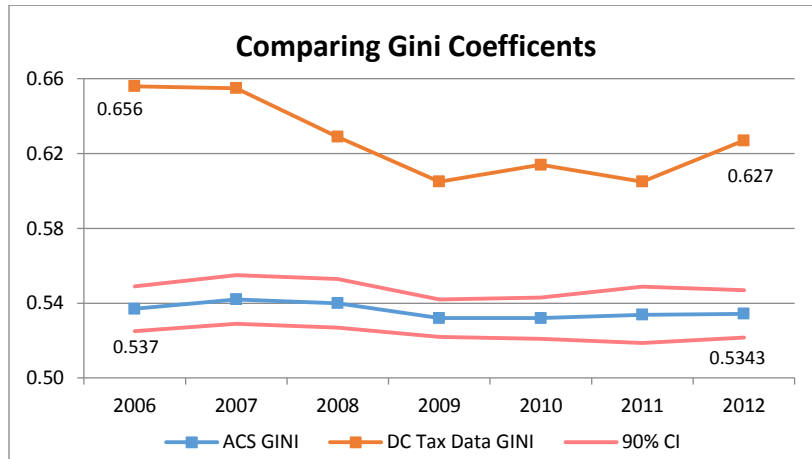
There are two main findings from the data presented in this section. First, compared to comparable statistics derived from local tax data, ACS income statistics are systematically and consistently higher and ACS Gini coefficients are systematically and consistently lower than comparable statistics derived from the tax data. And second, and maybe more importantly, some ACS and tax data reveal considerably different trends for the same phenomena. Notwithstanding the relatively no growth in real mean income per the tax data for years 2006-2012, mean income actually declined in the final time period per the tax data. But ACS data depicts mean income in the city growing at an annual average growth rate of 3.27 percent over the study period. Also, ACS Gini coefficients for the city show very little appreciable change over the 2006 to 2012 time period. The largest change in the ACS Gini coefficients exists between 2007 and 2009 when the index declined 1.9 percent. In contrast, tax data shows a 7.6 percent decline in the coefficients during that same time period but also a subsequent 3.6 percent increase between years 2011 and 2012.

Possible Explanations

There are a few possible explanations for the large systematic differences in income statistics produced from ACS data and from the DC tax data. ACS data is based on self-reported survey responses that may be less objective than income information provided on tax returns. Tax returns must be accompanied and supported by income documentation, and all tax returns are eligible for audit by OTR. Tax audits also come with the high probability of financial penalties for the tax filer if violations are found with respect to tax rules and regulations. Thus, data submitted on tax returns are subject to a greater degree of accountability and responsibility, making the quality of information in the ACS and tax data a possible issue.

However, once over 300,000 tax returns are submitted to OTR annually for administrative review and verification, the tax database, in the aggregate, may be subject to some temporary distortions due to administrative and automation processing. In tax administration, there are

issues such as late filings, quarterly filings, filings needing immediate review before further processing, duplicate administrative records, amended returns, overpayments, underpayments, and data entry/processing issues. These issues cause tax data to be far from perfect, particularly in the very near term. But because OTR and every tax filer have a strong incentive to work together to get each tax record 100 percent correct in a timely manner, there is a self-correction tendency for all records to be quite accurate with the passage of time. In contrast, the conditions that manage the ACS survey responses over time may be a little different.

How well the two data sources describe the city's income characteristics is also an issue. Tax filers that choose to submit tax returns and report all of their income are a biased sample of all income earning residents for a given jurisdiction. This is important because there may be a significant number of income earning residents that do not file tax returns and/or do not report all of their income. This pool of non-reported income by residents may represent a nontrivial number of households for the city and tax database. Consequently, the means, methods and processes of the ACS may better deal with this issue of capturing information from all households in the city regardless if they earn income, file tax returns, or report all of their income. ACS survey respondents are selected via a statistical methodology that makes them, in total, highly representative of the jurisdiction's total population regardless of circumstances. The sum estimation of income characteristics of the city may be more comprehensively described by the ACS.

But, the biggest contrast between the ACS and the tax data used in this analysis is the basic socioeconomic unit of analysis. For this analysis, households are the basic socioeconomic unit for the ACS, and individual income tax returns are the basic unit of analysis for the tax data. In the ACS, a household includes all the people who occupy a housing unit. A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied. The occupants may be a single family, one person living alone, two or more families living together, or any other group of related or unrelated people who share living arrangements. The individual income tax return tends to represent a single tax filing income earning resident (and their spouse and/or dependents if applicable). In the District of Columbia, tax filers tend to be either single individuals (single filers), single individuals and their dependents (head of household filers) or registered couples (married/domestic partners) that live in the same residence (along with their dependents if applicable).

This distinction between ACS households and tax data returns appears to be the most significant explanation in explaining the difference in income statistics from the two data sources. For example, Figure 6 shows that the city's mean and median income levels using the tax data are below the comparable ACS statistics.

Figure 6: Mean and Median Income: ACS versus Tax Data

The Mean & Median Income of the Two Data Distributions for the 2009-2013 Time Period
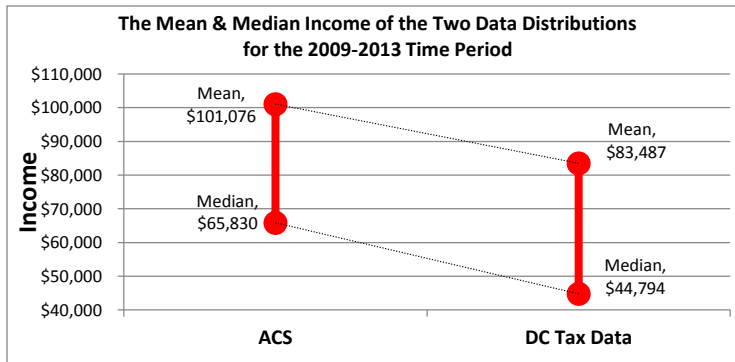
In terms of the means, Table 5 shows that there was an average of 323,253 annual tax returns in the city for the years 2009-2013. But, ACS data tells us that there was an annual average of 263,649 occupied housing units in the city for the same time period. Dividing the city's total income by the number of tax returns automatically yields a higher mean number than dividing the total income by the number of households because there were 22.6 percent more tax returns than households. In terms of the medians, the ACS tells us that the median income for households in the city was $65,830, but the tax data tells us that the median income for tax filers was $44,794. According to the ACS, the average household size in the city ranged between 2.11 and 2.31. With 60.9 percent of the city's tax filers being single residents, the tax data median of $44,794 most likely represents a single filer's income. And indeed, the panel of tax data for this analysis reveals that there were 22 tax records (tax filers) with an exact income of $44,794 for the 2009-2013 time period, 59 percent of them were single filers, 23 percent were head of households, and 14 percent were married filers. Consequently, the ACS median household income, which represents a household with an average of slightly more than two residents (and possibly two income earners), is largely being compared to an income of an single individual tax filer.

Table 5: Income by Tax Filer Type 2009-2013 (DC Tax Data)

	# Total Filers	% of Filers	Avg # of Filers Per Year	Income			
				Total ($millions)	% of Total	Mean	Median
Single	984,672	60.9%	196,934	$57.844.1	44.2%	$61,456.0	$44,038.0
Married	308,135	19.1%	61,627	$61.744.2	47.1%	$200,886	$114,929.0
Head of Household	281,014	17.4%	56,203	$114030.4	8.4%	$39,252.0	$28,052.0
All Others	42,443	2.6%	8,489	$360.6	0.3%	$9,187.0	$4,553.0
Total	1,616,264	100.0%	323,253	$130,979.2	100.0%	$83,487.0	$44,794.0

Given that the economic units being compared (households versus tax returns), may be a major cause for the discrepancy in the income statistics using the ACS and tax data, it is necessary to compare the overall income level of the city from the two data sources. Table 5 shows that the inflation–adjusted total income for the city for the five year period of 2009-2013 was $131.0 billion. The comparable income amount for the same time period from the ACS was $150.8 billion. Notwithstanding the estimated $19.8 billion (13.1 percent) discrepancy between the two

data sources, which may be an estimate primarily of the income earned by city residents that was not reported to tax authorities over the five year period, total income numbers for the time period under investigation from the two sources ($131 billion versus $151 billion) are fairly similar. Thus, controlling for the inability to classify income tax data by households, it appears that the tax data generally supports ACS income numbers for the District of Columbia in the aggregate.

A final issue raised in this analysis are the occasions where ACS and tax data points diverged appreciably as in the case of the differing mean growth trends in Table 4 and Gini coefficients (for year 2012) in Figure 5. Table 4 shows that while the mean income grew at an annual average rate of 3.27 percent over the study period, the tax data show mean income declined at an annual rate of 0.04 percent with the income in the final time period being less than the prior time period. This may stem from the facts that over 60 percent of the city's tax filers are single filers, and single filers have been the fastest growing cohort of individual income tax filers over the study period. According to the ACS, the number of occupied households grew at an annual average rate of 0.3 percent between years 2006 and 2013. According to the tax data, on the other hand, the number tax returns grew at an annual average rate of 2.3 percent over the same time period. And more strikingly, single filers grew at an annual average rate of 2.4 percent, signifying that the number of single filers grew faster than all other tax filer types. Furthermore, 61 percent of the total growth in tax filers from 2006 to 2013 was accounted for by single filers, and there were 9,332 (4.7 percent) more single filers in 2012 than in 2011. All of this suggests that the robust in-migration of single filers from 2006 to 2013 may have also been a major reason for the decline in average income per the tax data.

In sum, there has been a relatively large and continuous increase in new single filers migrating to the city. But, with single filers having a mean income 69.4 percent lower than married filers (married filers and single filers account for over 90 percent of all income), the interplay of these dynamics may be the cause of the Gini coefficient per the tax data for 2012 to trend upward while the ACS comparable statistic remained practically unchanged for years 2011 and 2012. For the study period, all of this suggests that while the out-migration of residents in the District of Columbia has been notable, the in-migration of new single residents has been slightly larger helping to cause the growth in the number of households to remain relatively unchanging. These important dynamics appear to have a significant bearing on the comparison of income statistics for the city when described in terms of tax filers and households.

Conclusion

Local jurisdictions sometimes maintain a vast store of administrative data that captures some of the same socioeconomic measures included in the ACS. This naturally leads to the question of how these measures from alternative data sources compare to one another. This analysis compared selected income related statistics from District of Columbia income tax returns to comparable statistics from the ACS.

This analysis finds great similarity in the citywide income level statistics as described by the ACS and DC tax data. But means and median statistics from the two data sources differ substantially and systematically likely because the tax data is often explained in terms of tax filers/tax returns and the ACS often explains income, more appropriately, in terms of households.

It also appears that the city's demographic changes, and consequent changing tax filer profile, contributed to some of the differences in the levels and trends in income statistics from the two data sources.

Without question, the ACS is a leading source of data on economic, employment, housing, demographic conditions and circumstances. However, this analysis underscores the importance of local leaders better understanding ACS data and methodologies and attempting to reconcile local and ACS data, particularly when local jurisdictions have similar, complimentary or supplemental data. Doing so could lead to a better understanding and better explanations of important changes occurring in local jurisdictions. Also, such an exercise could better inform many as to the strengths, weaknesses and the appropriate uses of each of these data sources.

Jayron Lashgari was a research intern with the Office of Revenue Analysis for the summer of 2015. She is currently an undergraduate Finance and Economics major in her junior year at Tulane University.