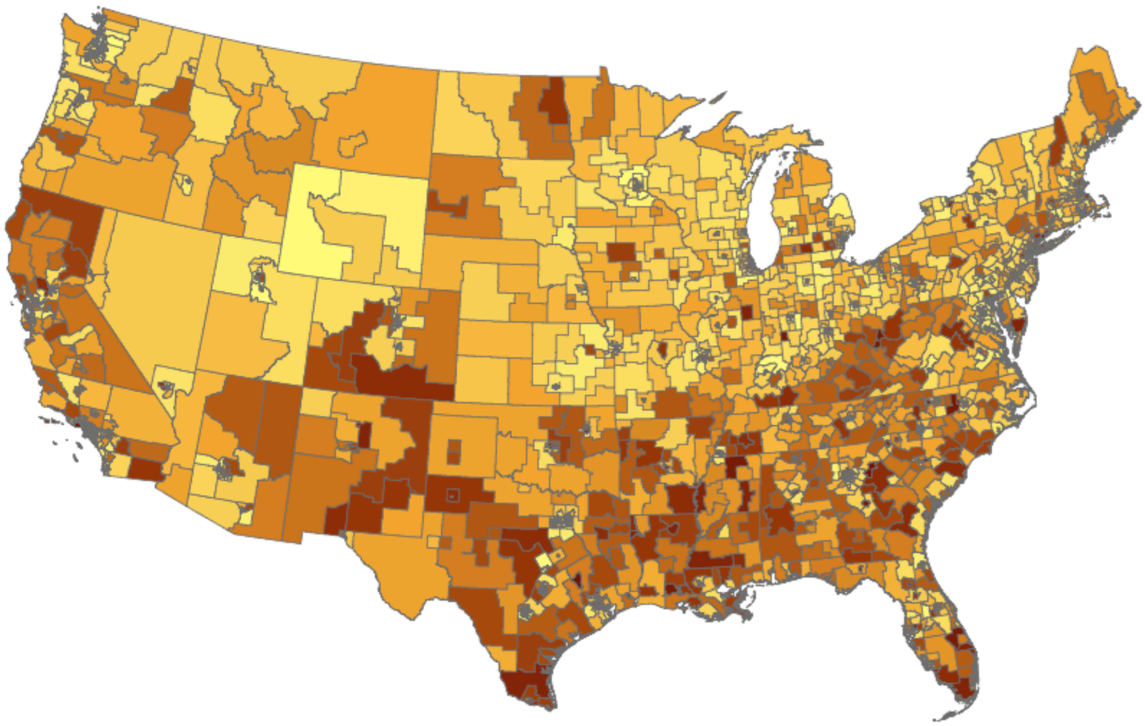

Effects of Geographic Aggregation on Inequality Indices

LSE MPA Dissertation 2015

Candidate #55291

Acknowledgement

I would like to thank Stephen Jenkins, Andrew Samwick, Dan Feenberg, Berkay Ozcan Sam Abrams, and my father for their assistance, guidance, and support.

Table of Contents

Table of Figures

Acronyms and Abbreviations

PUMA	Public Use Microdata Area
PUMS	Public Use Microdata Sample
ACS	American Community Survey
MAUP	Modifiable Areal Unit Problem

Abstract

This paper seeks to measure the effects of geographic aggregation on measures of American income inequality. This is done using Public Use Microdata Samples (PUMSs) from the American Community Survey (ACS). Annual measures of household and individual Gini and Theil indices are calculated. This is done at both the U.S. state and sub-state level. The distributions of these indices are compared and then mapped using GIS software. These maps are used to analyze the spatial distribution of income inequality in the United States. Several other statistics are calculated and mapped.

Evidence of both a strong effect of geographic aggregation as well as high spatial segregation is found. The distributions state and sub-state indices are distinctly different. Clear spatial patterns of inequality are discernible at both the national and metropolitan levels. This finding has important implications for the American income inequality literature. Furthermore, the method of index calculation provides a useful resource for future analyses.

Introduction

This paper seeks to measure the effects of geographic aggregation on measures of American income inequality. Given the recent focus on the rise American income inequality, as well as its causes and effects, obtaining accurate and useful measurements is crucial to the rigor of the scholarship effort. Before any causal linkages can be tested the integrity of the inequality measurements must be sound. This paper seeks to both measure and overcome what is termed a 'geographic aggregation effect' in estimating inequality.

First the relevant literature on American income inequality, spatial issues, and data sources for inequality is described. The presence of a geographic aggregation effect is then examined. This is accomplished by using a new method for obtaining measures of income inequality using microdata from the American Community Survey (ACS). The procedure combines several existing data processing methods to obtain geospatial estimates of local income inequality.

This provides several distinct advantages over other inequality data resources currently in use. By using individual records, a greater range of inequality indices is now readily available. Other commonly used sources for income data, such as the Current Population Survey, do not provide the level of detail of individual income in the Census data. These records also allow the calculation of indices at much smaller geographies than are previously available for annual estimates, as well as rough calculations of individual tax liabilities.

A principle role of this paper is to demonstrate the need to perform spatial analysis and the advantages of using rich new source of microdata. It explores the existing data sources used to measure geographic American income inequality and provides a useful alternative. It takes advantage of a new geography provided by the Census bureau to accompany the microdata files. This sub-state geography, known as a PUMA, provides an excellent alternative to larger levels of aggregation such as U.S. States.

To demonstrate the effectiveness of this new calculation method, a preliminary exploration of the geospatial aspects of American income inequality is performed. The distribution of the data is calculated and compared at different levels of aggregation. By examining the distributions of both PUMA and State inequality indices, it is found that there is a substantial difference between the two in terms of dispersion, smoothness, and median values. This is evidence of a geographic aggregation effect on estimates of income inequality.

The new estimates are then mapped and analyzed using GIS software. This is done at both a national level and for several select urban areas including San Francisco, New York, and Houston. A high degree of spatial influence is found for the vast majority of measures. This corroborates the findings of (Peters 2013) and others that American income inequality is highly spatially segregated.

The aggregation effect observed in the distribution comparison is corroborated by examining the difference in inequality estimates between PUMAs. Especially in populous urban areas, substantial differences between the state and PUMA estimates are visible. This shows the degree to which these areas would be 'mischaracterized' if they were analyzed using state-level estimates.

Other contributions of this paper are methodological. Current estimates of American income inequality have several weaknesses that this paper helps to alleviate. Many studies of inequality are performed with large geographic units that belie the immense diversity of place-based income inequality. Most other data sources for estimates of the income distribution are not microdata, meaning that substantial imputation may be involved. Using microdata allows a much more accurate calculation of the income distribution.

The analysis section of this paper seeks to investigate the spatial nature of the data. Several recurring patterns are observed, particularly with regards to urban areas. There is a common pattern of a dense, highly unequal urban core surrounded by a much less dense periphery with a high degree of income segregation. The relationship between geographic income shares and inequality is also examined, finding that increasing levels of income inequality is likely to be accompanied by a disproportionate share of the national household income.

Both the methodological and substantive contributions of this paper are enhanced by the inclusion of a method pioneered by (Samwick 2012) through which ACS data is paired with NBER's TAXSIM calculator. This yields estimates of total taxpayer income as well as federal and state tax liabilities. From these values post-tax inequality estimates are computed and mapped. The location-specific difference between these pre- and post- tax estimates is also mapped. This helps paint a picture of the mediating role taxes have in the spatial distribution of inequality.

This spatial analysis includes several measures not yet examined in the literature. The temporal variation in the inequality levels within each PUMA provide a picture as to precisely which areas experienced the greatest fluctuations in income inequality. This provides excellent direction for future investigation as to the causes and effects of these large, short term swings in inequality.

Literature Review

Inequality Literature

Concern regarding a rise in income inequality is a becoming defining feature of early 21st century politics and economics. Rising income inequality in many western countries beginning in the 1970's has spurred an extensive investigation of its causes, trajectory, effects, and potential solutions. The United States may be experiencing what some call a "great U-turn" in terms of income inequality. This refers to the relatively recent reversal of a decades-long gradual fall in income inequality that occurred during the postwar era (Peters 2013).

The effects of income inequality are perhaps even more contentious than its trajectory. A growing body of work seeks to substantiate the hypothesis that—through a combination of different processes—income inequality is persistently associated with increased societal ills, such as crime or various health related outcomes (see Wilkinson 2010). The association between health outcomes and income inequality is one of the more closely studied relationships thus far, with a growing number of studies supporting this association. However, there is less substantiation of a direct causal mechanism between inequality and poor health outcomes—the strongest such evidence is for 'psychosocial mechanisms" (Pickett 2015).

American income inequality is a politically contentious topic. Even among those that agree that extreme income inequality has pernicious or deleterious effects, there is little agreement about the policy options available. Some advocate for opportunity-related interventions, such as robust education systems, workforce development, increasing job benefits for low-income earners, and offering tax credits for working parents (University of Wisconsin 2015). Others such as (Piketty 2014) argue that stubbornly high income inequality is an inevitable feature of the later stages of industrial capitalism. This conclusion generally demands a more direct approach to addressing the growth in inequality through taxation. Yet for many in the United States, such aggressive government intervention is anathema to 'small government philosophies. There are scholars who question or dismiss the idea of tax or policy interventions, and even dispute the prevailing evidence that American income inequality has been steadily rising (see Perry 2014).

In the American case, urbanization appears to play a consistent role in income inequality. Urban areas are the sites of some of the most extreme values of income inequality, as they gather both the prosperous and destitute in a small geographic area. This is illustrated by (Peters 2013), who finds highly unequal areas to be sites of both high income and high human deprivation.

The data used in this study are restricted to the years 2005-2013. This encompasses the period known as the "Great Recession," the most dramatic downturn in the U.S. economy in six decades. (Thompson and Smeeding, 2012) provide a very complete and cogent analysis of the general impact of the Great Recession on wage and family income inequality and poverty that parallels this paper in several ways. They find a significant impact on wage income during this period resulting in heightened levels of wage inequality. This trend was driven by a divergence between the wage earners and the rest of the distribution. While this paper is not designed to examine the effect of the downturn, it shares many thematic and methodological areas with Thompson and Smeeding's report and does provide a valuable resource for further study of income inequality during this turbulent period.

(Thompson and Smeeding, 2012) also calculate the impact of taxes and transfers using NBER's TAXSIM calculator, which this analysis also uses. They find that taxes and transfers did help curb the sharp rise in inequality expected after a severe economic recession. However, this sheltering effect disappears when elderly households are excluded from the calculations. Finally, they examine the impact of household aggregation on wage inequality, finding that the Great Recession likely caused a slowdown in the fall of household sizes, which is a decades-long trend. This paper instead investigates the impact that household aggregation has on the calculation of inequality indices themselves.

The data analysis techniques provided by this paper hope to help clarify many of these debates. By providing annual estimates of the inequality across a range of different income sources, any ambiguity as to the precise trends will hopefully be alleviated. Thanks to the richness of the ACS microdata, a much broader set of income sources can now be analyzed. These yearly estimates can also prove valuable to investigating more causal questions such as the relationship between inequality and health outcomes, provided the PUMA geographies can be linked with health data. Urban areas are especially well covered by PUMA geographies and hold high potential for such analyses.

Data Literature

Existing analyses of American income inequality share several common data sources, including the ACS, the Current Population Survey, and the Survey of Consumer Finances. This paper contributes a new method for calculating geographic income inequality that offers increased accuracy, flexibility when compared to previous sources. As a tradeoff, this new method sacrifices an extended time horizon and also requires the use of very large datasets.

While geography has been an implicit feature of the inequality literature, only recently have efforts been made to calculate, map, and analyze it in a more complete context. Few studies attempt to calculate income inequality for geographies at the sub-state level. Even fewer do so across multiple time periods.

Many of these studies rely on ACS Census data, the same principal data source as this paper, but they differ in several key ways.

(Peters 2013) performs a similar analysis, obtaining longitudinal data from the US Census and ACS data from 1970 to 2010. While he is successful in covering a much longer time period, this comes at a cost, as individual-level data was not available until recently. Thus any analysis of local income inequality that hopes to cover this time period or earlier must rely on a rougher approximation of where the population falls within income 'bands'. Because county-level data is not available at the individual level, (Peters 2013) and others must approximate the distribution of income from the number of households falling within a $10,000 income band. This is problematic, as it drastically reduces the resolution at which the cumulative income curve can be estimated. No information is available about the distribution of incomes within these bands, so additional assumptions must be undertaken that hinders the accuracy and comparability of estimates derived in this manner.

Several studies find a high degree of inequality across space, which this paper corroborates. (University of Wisconsin 2015) find a positive relationship between income inequality and adverse health outcomes using an approach to the spatial question that demonstrates both the advantages and tradeoffs of the techniques outlined in this paper. Their analysis uses a much blunter inequality measure, but also uses a generally smaller geography that is much more easily combined with other data sources.

This is one of the central tradeoffs that must be made when working with geographic ACS data: the availability of geographies and the compatibility with other data sources. The (University of Wisconsin 2015) analysis is possible because it uses an income inequality measure that is available at the county level. The 80/20 ratio is accompanied by all the methodological drawbacks of percentile ratios. Percentile ratios, which report the ratio between incomes at two different points of the distribution, are limited in that they simply use two points to summarize the entire income distribution, and thus may fail to report some distributional changes. Yet because these measures are so simple they can be released at a smaller geography without compromising confidentiality requirements. However, counties are a much more common geography in non-Census datasets than PUMAs. Depending on the inquiry it may still be necessary to use more restrictive measures in order to use data that is geographically comparable.

By using the microdata now available from the ACS, this paper fills a gap in the literature by describing a method by which annual estimates of inequality indices can be calculated at the sub-state level. This approach has several advantages over two other common methods, including using official Census statistics and deriving results from a survey such as the CPS.

Since 2006 the ACS has calculated and released Gini estimates at the national, state, and sub-state level. For some research purposes this may be adequate, but there are several drawbacks. Annual data is not available for many of the smallest geographies, meaning that one has to make a choice between obtaining yearly estimates and using a favorable geographic level.

Consistent calculation of a wider variety of inequality indices will provide a much more stable and flexible resource for estimating the effects of policies and events on income inequality. It will also allow better exploration of the causal relationship between income inequality and various social, political, and health outcomes.

Spatial Literature

Geographic space is an important feature of income inequality. As income inequality has increased in the United States, it has not done so uniformly across geographic space. (Peters 2012).

Most spatial analyses of inequality use US states as the geographic unit (Peters 2002). When attempting to draw conclusions related to specific policies this is a logical unit of analysis, as US states have a wide latitude in crafting their tax, welfare and education policies. In this case, state borders serve as effective and credible boundaries for investigating the relationship between inequality with specific state-level features. Some analyses do attempt to use or report inequality estimates at a sub-state level. However attempting to do so, as this paper does, usually means sacrificing precision in terms of estimates. It is also uncommon to be able to calculate alternative measures of inequality such as the Atkinson or Generalized Entropy indices, as these are less commonly reported by official bodies and require individual income data to compute. Providing a way to easily generate these less-common indices and map them onto small geographic units is a chief contribution of this paper.

Modifiable Areal Unit Problem.

The Modifiable Areal Unit Problem, or MAUP, is a longstanding issue is spatial geography and can pose a severe hindrance to geographic research. MAUP refers to the potentially corrosive effects that spatial aggregation can have on population estimates and aggregated statistics.

MAUP has two components, both of which can affect estimates. Firstly, the variation in the size of geographic units has been shown to sometimes drastically affect even simple correlation estimates (Openshaw 1979). This is known as the 'scale effect'. The second effect comes from the division or zoning of geographic units within a particular area. This is particularly problematic when it comes to estimating societal effects or associations with income inequality. For the most part, neither the size nor the shape of geospatial units is easily manipulated. Both income and geospatial data are treated as highly confidential, meaning that researchers are generally stuck with a single unit of analysis.

There is no standard method for overcoming the MAUP, which has led to its long tenure as a problem for spatial analyses. Demonstrating robustness to the MAUP is difficult and demands extremely detailed spatial data. (Fotheringham 2000) recommend varying both the size and shape of spatial units to demonstrate robustness, but this requires a level of detail at the individual level that may not be available. While (Bryant et al. 2010) and others have attempted more standardized methods of demonstrating robustness, the MAUP remains an open question in the spatial literature. Calculation of measures at different geographies is but a small step in addressing this complex problem.

Data and Methods

Because of the static nature of PUMA delineations, investigation of a zoning effect is not readily possible. While the ACS data used in this paper are not detailed enough to exhaustively investigate the MAUP, the measurement of inequality at the sub-state level does provide an initial measurement as to the presence and severity of a scale effect.

This paper reports a new method for calculating income inequality estimates. Establishing a stable, reproducible, and effective method for generating inequality measures from this income microdata paves the way for closer and more rigorous examination of the relationships between inequality and other societal factors.

In order to investigate the effect of geographic aggregation on measures of inequality the PUMS data is combined with PUMA shapefiles. Both are described here.

The Public Use Microdata Area

This analysis uses the Public Use Microdata Area, or PUMA. This is the smallest geography available for use with the ACS PUMSs. Because I intended to show the full capability and limitations of using PUMS data for geographic analysis, a detailed description of PUMAs, including their delineation, common traits, and applicability to other data sources, is necessary.

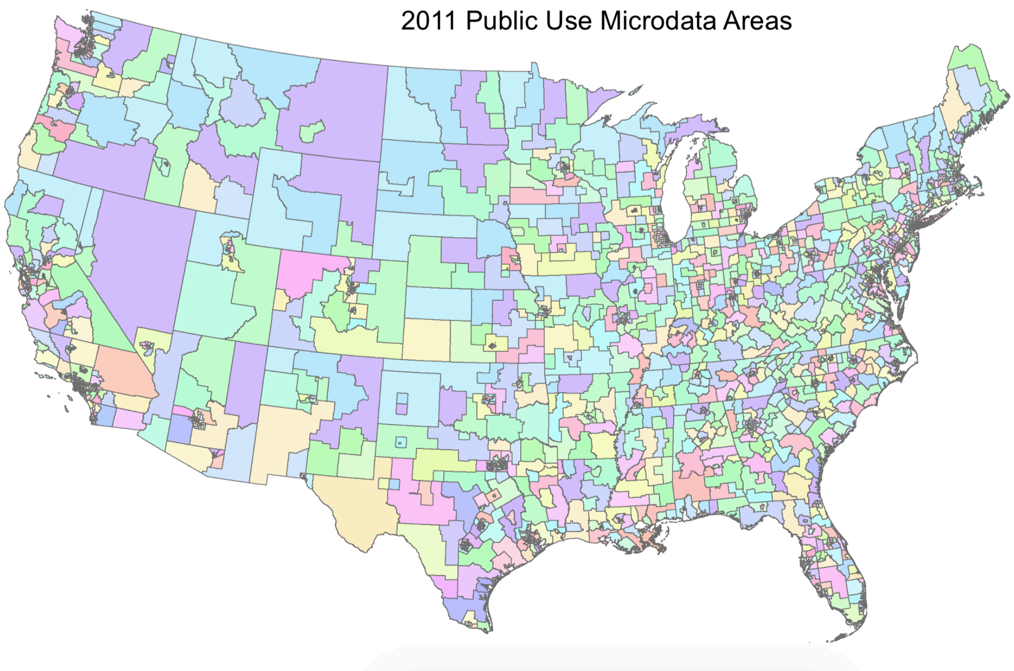

FIGURE 1

PUMAs are a new geography initially defined in the release of the microdata samples from the 2000 census. In 2005 the ACS began releasing PUMS data coded within both states and PUMAs.

The Census Bureau provides requirements for the delineation of PUMAs. For the 2000 PUMAs, which this paper primarily uses, PUMA boundaries were designed to correspond to 5-percent samples of the master ACS and Census data. PUMAs must contain at least 100,000 people, which means that they vary in size from small sections of a county to large swaths of a state. Figure 1 shows the 2000 PUMA delineations for the entire country. Many of the smallest urban PUMAS are not visible, but clusters of them are readily apparent.

While the geographic area can vary substantially, the population counts of the 2070 pumas are relatively consistent. PUMAs must be contiguous within state lines. The boundaries can only follow counties, census tracts, and a handful of less-common entities. This equivalency is important, as the PUMA is a relatively new and uncommon geography outside of ACS and Census data. Although ACS PUMS files available for 2001-2004 contain a ‘puma’ variable, this variable in fact only refers to the states (Census 2005). It is important to note that while PUMAs are uniquely identifiable within a state, they are not uniquely identifiable at the country level. In order to reproduce the national PUMA results in this report, new identifiers must be generated. Multiyear ACS estimates do contain geographies smaller than a PUMA, although those are not used here.

PUMA Adjustments

The PUMAs were recalculated and redrawn after the 2010 census according to a slightly modified set of delineation guidelines. These new PUMA delineations took effect with the release of the 2012 ACS data. This analysis uses data from both PUMA vintages. All maps are drawn using 2010 or 2011 PUMA data, which are drawn within the 2000 vintage. Where full distributions are calculated, data from the 2012 and 2013 PUMS are included.

Although PUMA redefinition generally occurs only after a new decennial census, there are certain circumstances where this is not the case. One such adjustment occurred between 2005 and 2006. Due to population displacement from Hurricane Katrina, three PUMAs in Louisiana no longer met the 100,000-person population requirement. Thus there are effectively three current PUMA vintages, not two. In order to make the data comparable for this analysis, these Louisiana PUMAs #01801, #01802, and #1905,in the 2005 PUMS have been merged into their parent PUMA, #077777 (Census 2007).

Working with PUMA geographies entails paying particularly close attention to the nuances described here. While the Census releases detailed documentation for the PUMS data, there are several key adjustments that must be made before any inequality index can be calculated.

ACS PUMS Data

The inclusion of PUMA geography in the release of the 2005 ACS PUMS data marks the earliest time period present in this paper. PUMS, or Public Use Microdata Samples, are a subsample of the yearly ACS and contains responses for almost every item asked in the ACS. PUMS data is also released with the decennial census, but that is not used here. The PUMS Data has two components, the individual and the household record. Each individual record corresponds to a 'parent' household record. (Census 2001). The ACS is a representative survey of all U.S. Households and group quarters, so households do include non-related individuals and families. This paper relies on PUMS data from 2005 forward, when the ACS switched to a constant sample of 250,000 addresses per month. Interview dates for the microdata observations are not provided.

Income Definition and Comparability

The ACS reports different income sources at the individual and household levels. At the household level two income aggregations are available, family and household. Because households may include income from nonfamily members, this measure is usually larger. The Household record contains pooled data for family and household income, whereas the individual record contains measures on a wide variety of income sources such as wages, retirement income, and public assistance. This paper primarily uses the total household income variable, 'hincp'.

There are both top- and bottom- coding issues in the PUMS data of which to take note. There are many observations that report zero or negative values for income, even for total personal or household income. When the Census Bureau calculated Gini coefficients of household income it recodes these observations as zero and includes them. The Stata command used to generate measures of income inequality in this paper, *ineqdeco*, simply omits these observations.

It is also important to note that the ACS PUMS microdata are not necessarily comparable to income measures in other surveys. Although they are highly similar, income definitions between the ACS and other Census-administered surveys such as the CPS or the decennial census differ enough to not be directly comparable. The decennial census routinely reports slightly higher income values than the ACS. This may be due to a difference in reference period--while the decennial census asks users to report their income in the previous calendar year, the ACS asks them to report it over the previous 12 months (Census 2009).

Inflation Adjustment

As this paper is primarily concerned with calculating measures of within-year inequality, no adjustment has been made to put between-year incomes surveyed in constant dollars. However, within-year adjustments must be made each year. ACS PUMS income data are reported in nominal dollars and must be adjusted by multiplying the value by the 'adjust' variable for years 2005-2007 and the 'adjinc' variable for years 2008-onwards. This is done because the ACS interviews occur throughout the year—the date of interview is not reported. As income is reported for the 12 months preceding the interview, Derived from BLS's monthly Current Price Index Using Current Methods (CPI-U-US) this adjustment places all incomes in constant July dollars of the survey year (Census 2009). This adjustment is not necessary if one is using three or five year income measurements from the ACS.

Inequality Measures Calculated

Once appropriate adjustments have been made, various income inequality measures are calculated using the Stata commands *ineqdeco* and *inqdec0* (Jenkins 2008). Although this paper focuses on the Gini and Theil indices, the microdata can easily be used to calculate others such as percentile ratios, Atkinson indices, or the Mean Logarithmic Deviation.

Because of the aforementioned top- and bottom- coding issues in the ACS data, this analysis will primarily use middle-sensitive measures such as the Gini coefficient and the Theil T, or GE(1) index. The assumed accuracy of the middle-distribution data also discourages the use of percentile ratios, which can ignore changes in this part of the distribution. The Theil index is one of the two middle-sensitive generalized entropy indices. It begins at zero and increases in value as inequality increases (Jenkins 2009).

The Gini coefficient is the most commonly used inequality index and is used for the majority of this paper. It ranges from zero to one, and measures the given distribution's 'distance' from perfect income equality, with higher values indicating more inequality (Jenkins 2009). These estimates will differ from the Census-released Gini coefficients. The bottom-coding procedures are slightly different, and the ACS estimates are calculated using the full ACS sample from that year, whereas the PUMS is a 5% sample of this data (Census 2009).

These indices are then used to investigate the geographic aggregation problem in several ways. Firstly, the distributions of PUMA and state index values are compared to see if there is a substantive difference in their shape and values. These indexes are then shaded on a map of the US PUMAs by quintile to see if there are any apparent spatial patterns. Finally, in order to gain a better understanding of the features of the spatial component of income inequality, several other statistics are used.

Population Share and Income hare

The population share is calculated for both household and individuals at the PUMA and state levels. For this I use the Stata package *ineqdec0*, a command closely related to *ineqdeco* that allows for the inclusion of income values that are negative or zero (Jenkins 2008).

Also calculated is the percentage share of national household income that is encompassed within a particular PUMA.

Post-tax Estimates: Using TAXSIM

Calculating post-tax estimates from PUMS data is tricky, but (Samwick 2012) develops a way to make PUMS data compatible with NBER's TAXSIM calculator, which I will summarize here. TAXSIM is a web-based application designed to simulate the federal and state tax liabilities for a theoretical taxpayer. It treats each observation as a single tax return, requiring a set of inputs that make up Adjusted Gross Income, including capital gains, wage and public assistance income.

TAXSIM produces highly accurate estimates of federal tax liabilities and marginal rates, whereas the state liabilities are less accurate (Feenberg 1993).

This paper uses the adjustment procedures to make ACS PUMS data compatible with TAXSIM developed by (Samwick 2012). While this allows the use of resource, there are some limitations. The ACS survey does not have data for a number of components of the Adjusted Gross Income that tax liabilities are based on. Some of these, such as charitable donations, are imputed using IRS statistics of income and year. Other components, such as childcare expenses and unemployment compensation, are set to zero.

These post tax estimates of inequality have strengths and weaknesses. Post-tax estimates of inequality give a much more accurate picture of the income that households actually receive. Thus these levels of income—and income inequality—reflect income values that are more likely to correspond to income-related outcomes. The inclusion of state incomes is particularly valuable, as it accounts for another spatial factor in income inequality—state tax schemes. However, due to the somewhat awkward fit between the ACS data and the requirements of TAXSIM, these estimates are secondary to the central analysis of this paper.

Results and Analysis

This section explores and analyzes the data in several stages. First the distributions of inequality indices are examined. These indices are then mapped by state and PUMA, showing the spatial distribution of income inequality. Finally, other measures relating to these the indices are also mapped and analyzed. In instances where a single vintage is studied the 2000 PUMAs are used, as more comparative data is currently available.

The Distribution of Inequality across PUMAs and States

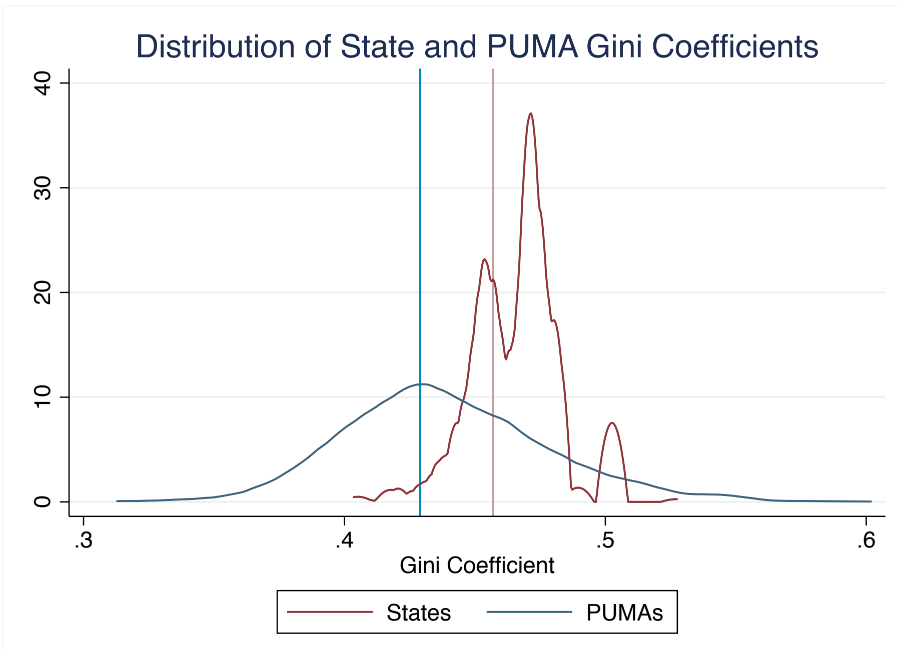

FIGURE 2

Figure 2 shows the kernel density distributions of the Gini coefficients of U.S. states and PUMAs for 2005-2013. Since PUMA distribution contains all data 2005-2013, meaning both PUMA vintages are included. Both PUMA vintages are included in the distribution. Vertical lines are used to indicate the median values for both geographies. The PUMA distribution is much smoother—likely a function of the much larger number of observations in the sample. It is also much flatter and wider, indicating a much larger dispersion of values. Prior to any mapping, it is apparent that geographic aggregation at the state level drastically affects the observed measures of income inequality.

Figure 2 concisely illustrates a central problem with using state-level inequality measures. Despite their much wider availability, state-level measures will sacrifice a large amount of the variance present in the relative locations of income inequality. The aggregation that occurs between PUMA and State has significantly altered the data, despite covering the same total geography. From Figure 2 it appears that geographic areas towards the bottom of the income inequality distribution may be more likely to be 'smoothed over' by state-level aggregation. While no state has a Gini value below .4, a significant tail of the PUMA distribution falls below this value.

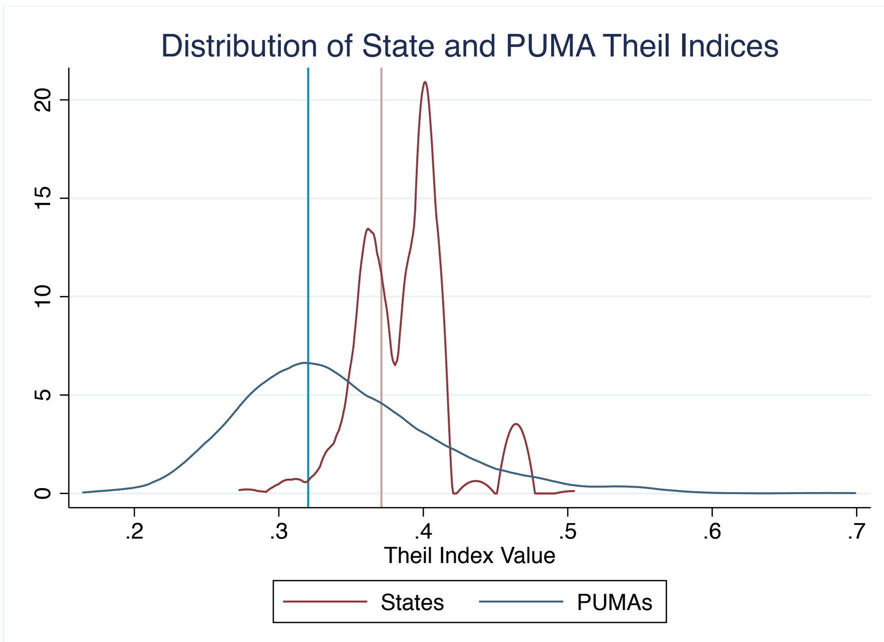

FIGURE 3

Figure 3 shows the same distribution comparison as Figure 2 with the Theil index. Here again the distribution of state Theils is further to the right and less smooth. The median value for states is again noticeably higher than it is for PUMAs. The PUMA distribution of the Theil index exhibits a much longer upper tail than the Gini distribution.

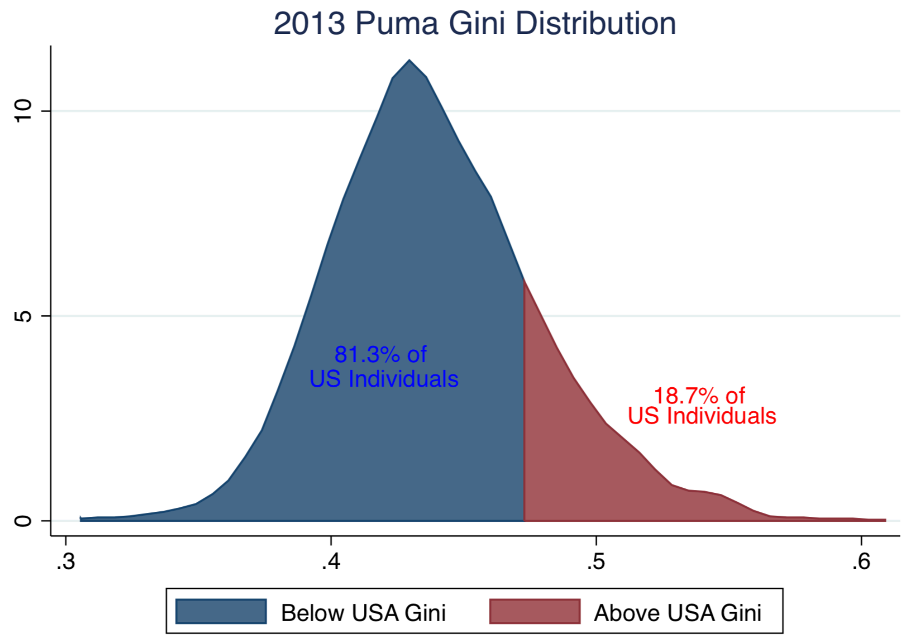

FIGURE 4

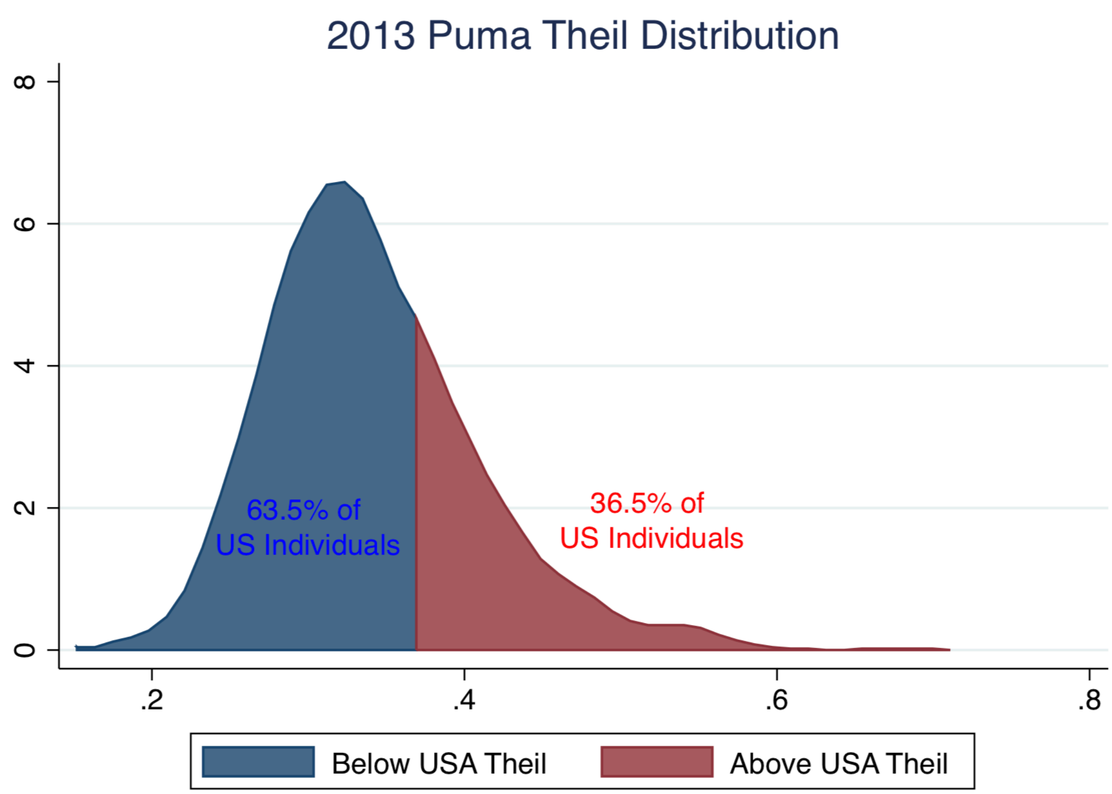

FIGURE 5

Figures 4 and 5 further illustrate the same issue with analyzing income inequality at large geographic levels. In each plot the kernel density distribution of the respective 2013 PUMA inequality index is broken into two areas. The dividing line between the upper and lower area is the 2013 inequality index value for the entire United States. Included is the percentage of American individuals age 18 or older that fall on either side of the line. While the density distributions reflect the PUMAs on either side of the USA index, the percentages reflect the individuals living in the PUMAs that fall above and below the USA index value.

Again the Thiel distribution shows a longer and flatter upper tail than the Gini. Additionally, the percentage of individuals residing in PUMAs in this upper tail is also much larger—36% vs. 19% —nearly twice as large. The difference in these two illustrates another important point—the importance of calculating and comparing multiple measures of income inequality. While the Gini index is by far more popular, it is by no means universal. Using ACS microdata allows reliable and reproducible calculation of virtually every measure of income inequality. This allows for much easier direct comparisons between different inequality measures such as this one.

TABLE 1

Gini Quintile (Household)	Top Gini Value (Household)	Population Share (Individual)	Income Share (Household)
1	.407	.19	.186
2	.427	.199	.193
3	.445	.203	.194
4	.470	.201	.198
5	.602	.205	.23

As a final examination of the dispersion of income inequality, Table 1 reports the population and income shares of each Gini quintile of the 2013 PUMAs. The individual population shares of each quintile appear relatively even. The largest quintile by population is the fifth, which has only 8% more than the smallest, which is the first. This is evidence that in the aggregate PUMAs provide a relatively even distribution of the U.S Population. However, the income share of the quintiles is much more uneven. Each quintile has a higher income share than the previous—as inequality increases, so does income share. As such, the 5th percentile has an income share 24% higher than the first.

This finding corroborates (Peters 2013), who found that U.S. geographic areas with less income inequality also tended to be more marginalized and impoverished. While it cannot be conclusively said from this table, it is certainly possible that the spatial clustering of higher-income households lead to both an elevated level of income inequality as well as an outsized share of the total national household income. This would explain both the even population shares as well as the steadily increasing income shares of each PUMA Gini quintile.

Mapping Inequality

In this section, both Theil and Gini coefficients are mapped and presented by quintile. With the exception of the mapped TAXSIM data, which is from 2010, all maps in this section are made with estimates from 2011 income data. Although PUMAs in Alaska and Hawaii are not included in any of the maps they are included in the tabulation of the index quintiles.

Figures 6 and 7 compare the spatial effect of aggregation on the calculation of Gini coefficients.

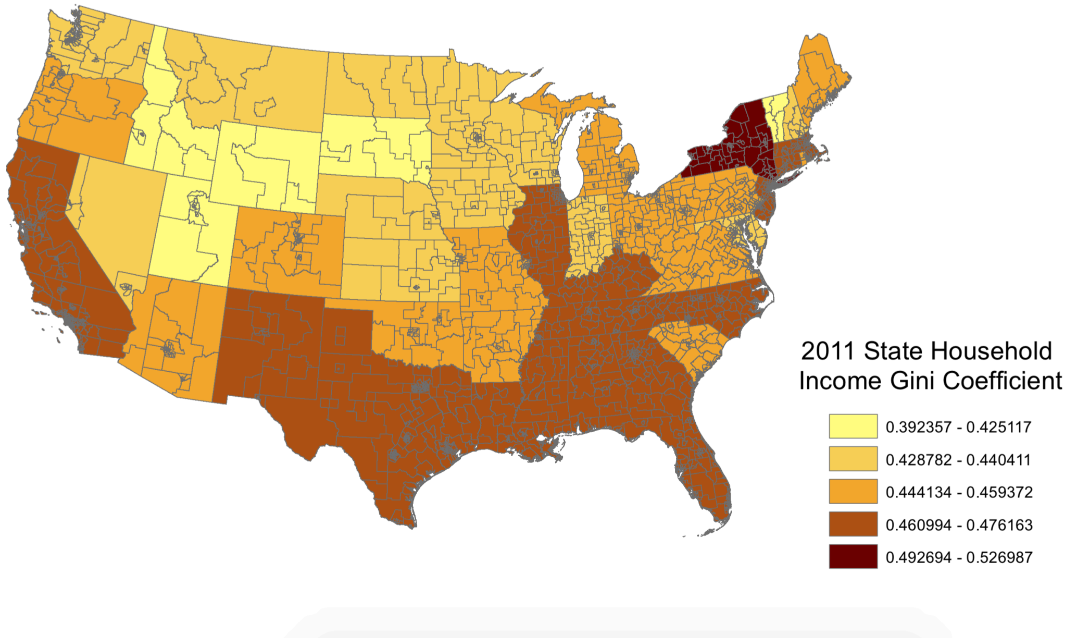

FIGURE 6

Figure 6 presents the state level household Ginis for 2011. From this map we can see a spatial distribution. The South and California. Much of the Midwest and Northwest appear to be more equal than the rest of the country. There are two clusters of extremely low income equality, one in the mountain west and one in the northeast. New York is the most visible state with relatively high income inequality.

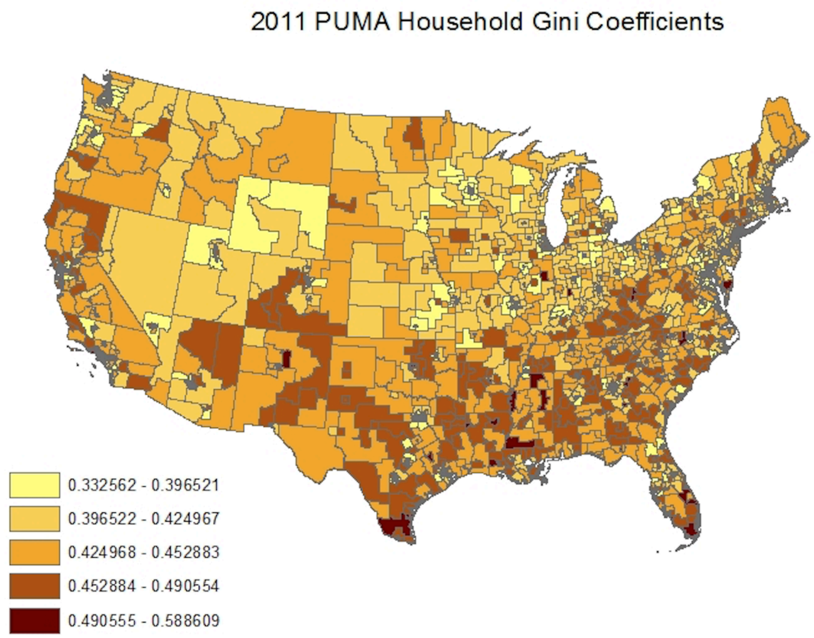

Figure 7

Figure 7 presents the same index calculated at the PUMA level. A much more varied spatial distribution is immediately apparent. California no longer appears to be an area of high geographic income inequality. It is likely that the inequality previously observed in Figure 6 across California is in fact clustered in the coastal metropolitan areas not visible at this scale. Large areas of the Midwest and Mountain West still appear to have relatively low levels of inequality in general, although PUMAs on the upper end of the inequality distribution are now visible in these areas.

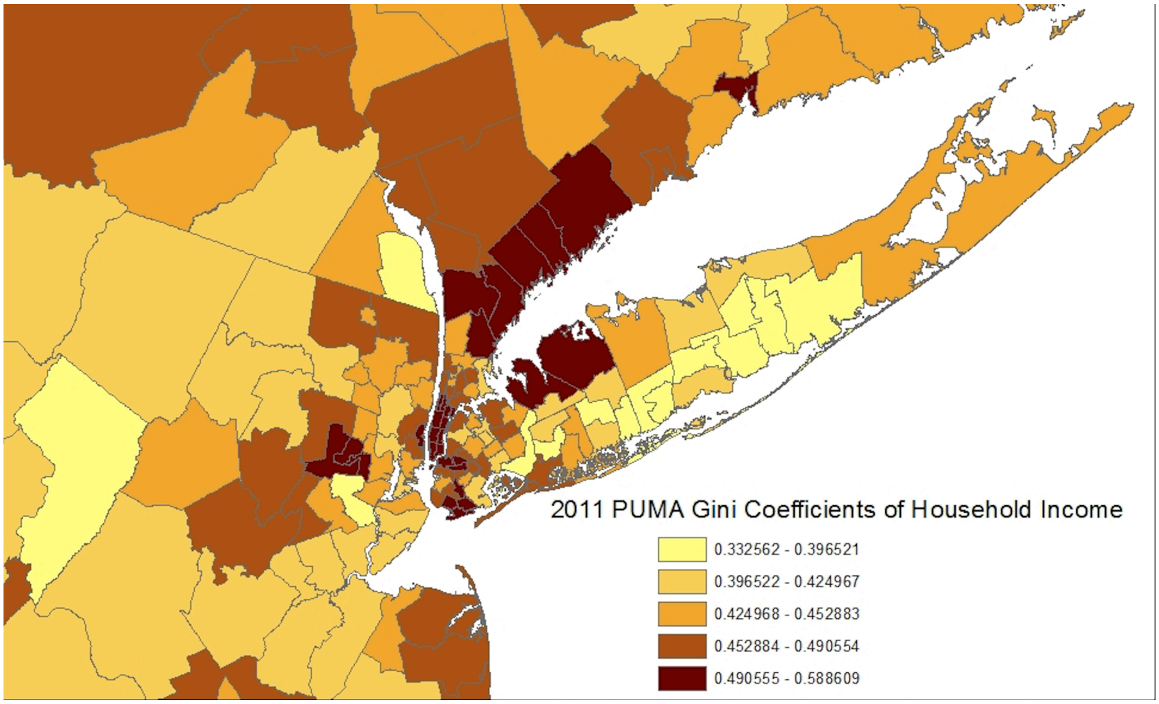

FIGURE 8

Figure 8 presents a closer look at the 2011 Household Gini for New York City and the surrounding PUMAs. Here the urban component of income inequality is immediately apparent. All of Manhattan lies in the upper quintile of the national PUMA Gini distribution, as does large parts of Westchester. Yet within a relatively short geographic distance lie many PUMAs on the other end of the inequality spectrum. These contiguous areas of income equality lie outside the city, indicative of a much higher level of income segregation outside the dense city center. The size of the PUMAs is also worth noting here. Due to the extreme population density of these areas, multiple PUMAs fit within a single county or city. This is in stark contrast less populous states visible in Figure 7 such as Nevada, which has an overwhelming portion of its geographic area encompassed by a single PUMA.

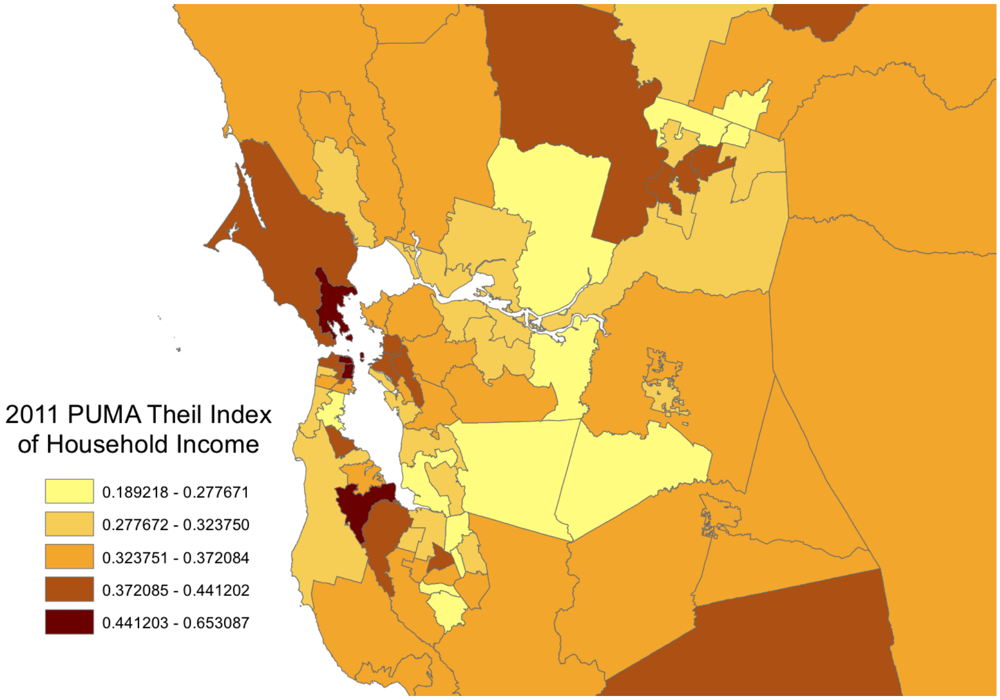

FIGURE 9

Figure 9 presents another look at urban inequality, this time in the San Francisco Bay Area and using the Theil index. The high levels of inequality in San Francisco and San Jose are contrasted sharply by low levels just across the bay in Fremont and Pleasanton. This pattern is consistent with findings in other studies such as (Weinberg 2011) and (Wheeler 2008) that have found increasing levels of neighborhood income segregation.

Post Tax Measures

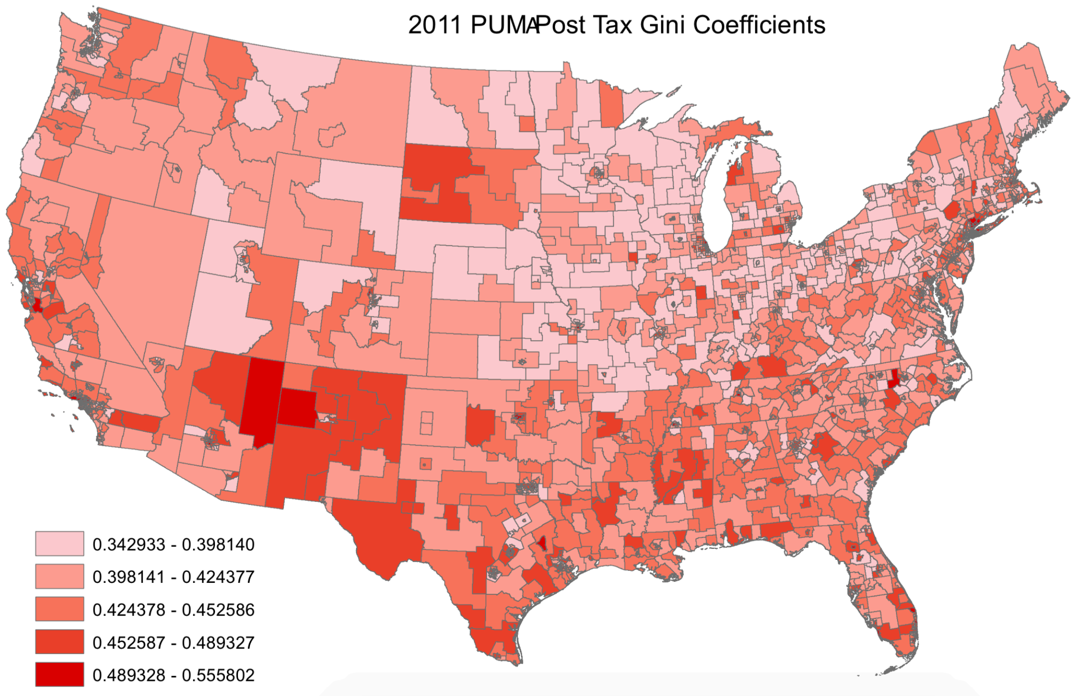

FIGURE 10

Figure 10 presents the 2011 Gini coefficients calculated from the post-taxes and transfer estimates from the TAXSIM model. Here the Midwest again appears to be relatively more equal. Southwestern areas such as Arizona, New Mexico, and southern Texas now stand out as places where income inequality is much higher and more widespread than the rest of the nation. Likely because state taxes are included in the TAXSIM calculation, individual states now appear much more distinct. South Dakota, for example is clearly visible, especially when compared to Kansas, its distinctly egalitarian southern neighbor.

Mapping Other Measures

This final section presents a further look at the inequality measures by generating and mapping a set of derived measures. Data is again from 2011 mapped onto the modified 2000 PUMA vintage, except for the standard deviation map, which displays data from 2005-2011.

Household vs. Individual

Because the ACS PUMS contains data on individual income, household income and household membership, it is possible to calculate inequality indices for the incomes of each of these units. Once this is done, the differences between them can be mapped in the same way as the indices.

In effect, computing income inequality indices at the household level—as is commonly done—is a form of aggregation. This aggregation is generally justifiable, as households generally function as a single unit in terms of taxation and consumption.

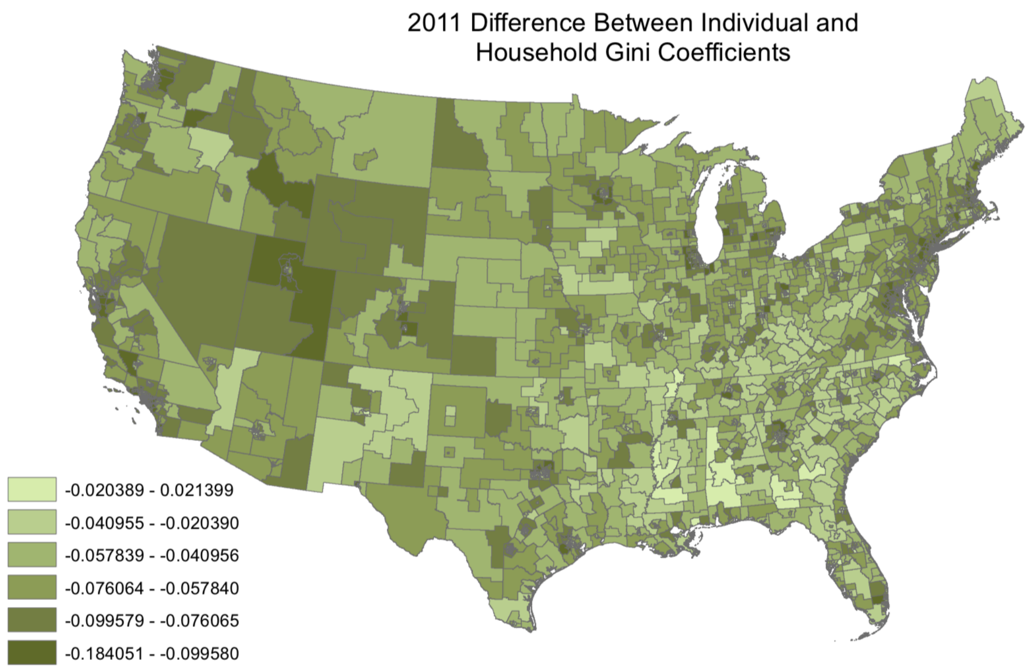

FIGURE 11

Figure 11 shows the effect of this individual-to household aggregation. It presents the difference between them Gini coefficients by subtracting the individual Gini coefficient from the household Gini for each PUMA. As individual incomes are inherently more dispersed than household incomes, individual Gini values are generally higher. This yields negative difference values for every quintile except the first. The darker areas of Figure 11 indicate areas where the clustering of individuals into households has effectively reduced the level of inequality. Again, this conclusion assumes that income is effectively dispersed within households evenly. Dark PUMAs are scattered across the country, with the only discernable large cluster visible along the Boston-New York-Washington DC Corridor. Also known as the 'Northeast Megalopolis,' this area has been the most heavily urbanized tract of land in the nation for decades (Gottman 1957).

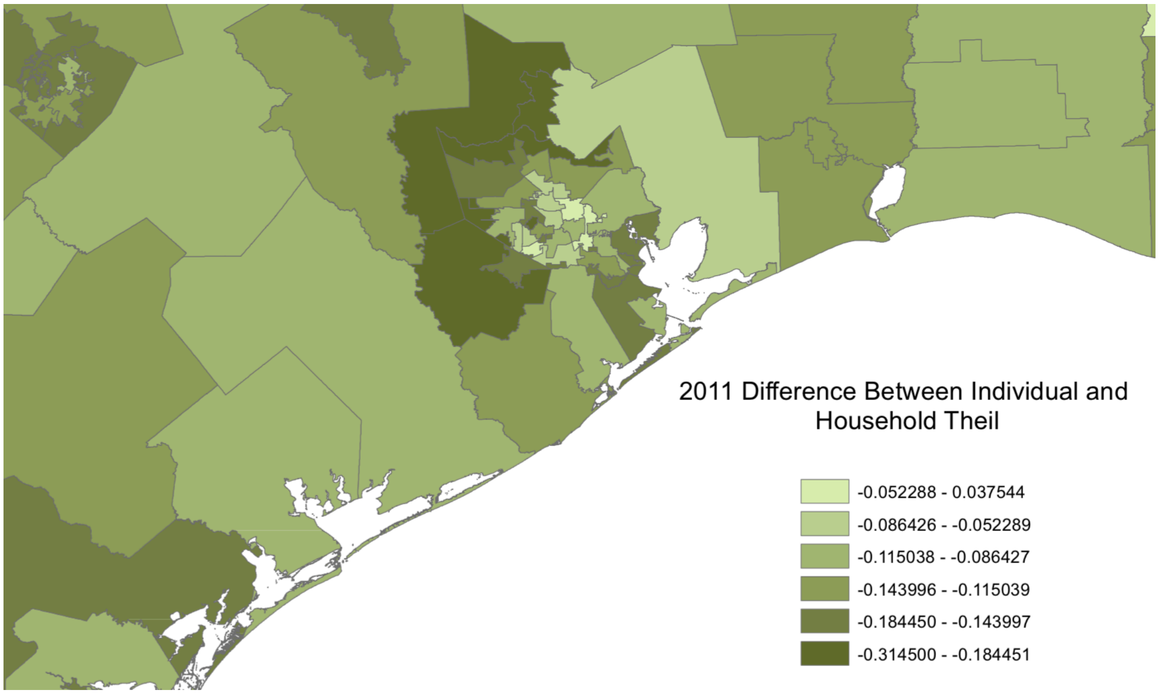

FIGURE 12

Figure 12 shows a closer look at the Individual-Household index difference for PUMAs in the Houston metropolitan area. This map, colored using the Theil index, shows the relatively higher proportion of multi-person households in the less dense suburban relative to the urban core. The smaller, lighter PUMAs encompass the denser metropolitan core. Further out housing is much less dense, resulting in larger PUMAs. These less dense areas seem to have a higher rate of aggregation of individuals into households, as indicated by the much larger reduction in the Theil value in these areas.

Clearly the choice of particular income source is of crucial importance. Indices calculated from different income sources will likely have a nonrandom geographic bias, and will thus be unsuitable for direct comparison. The choice of income source depends on the particular research interest. Household income provides a sound basis for measures of the socio-economic unit. However, to examine inequality directly related to labor market outcomes, measures derived from individual or wage income may provide more empirically sound estimates.

Effects of geographic aggregation

Figures 13 and 14 display the difference between the Gini coefficients of each individual PUMA and its parent state. These differences are calculated by subtracting the PUMA Gini from the state Gini value. As shown in Figure 2 state Ginis values tend to be larger than Puma Ginis, so a majority of the differences in this map are positive. However, because the distributions overlap, values are both positive and negative.

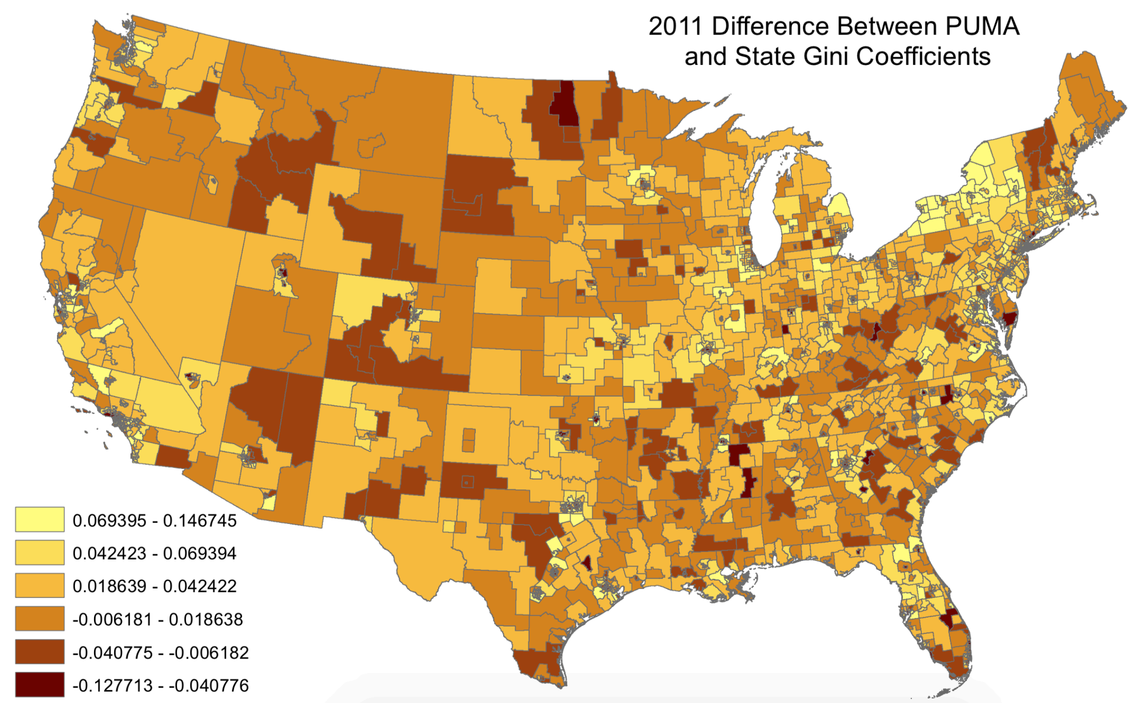

FIGURE 13

There appears to be little regional patterns in the differences in Figure 13, which makes sense as this particular measure is calculated within each state.

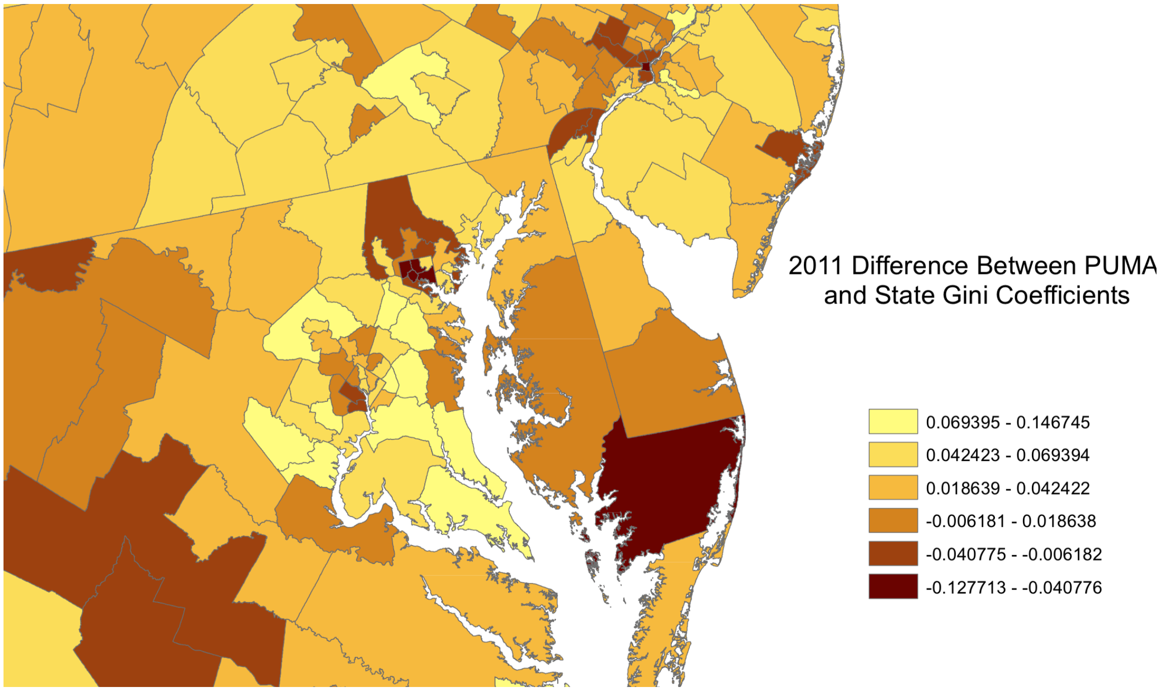

FIGURE 14

Figure 14 shows the Puma/State Gini difference values for the Washington D.C. metropolitan area. The dense urban core has a cluster of negative difference values, indicating that this area has a higher inequality than the state as a whole. The D.C. periphery lies on the other end of the spectrum, its positive values indicating that these PUMAs are well below the state Gini. This map essentially shows the degree to which both these areas would be mischaracterized in terms of income inequality if they were studied only using the state-level Ginis.

Income Share

Figure 15 shows a detailed look at the of national household income quintiles membership of PUMAs in the New York area. These income shares are calculated in an identical manner to the income shares reported in Table 1.

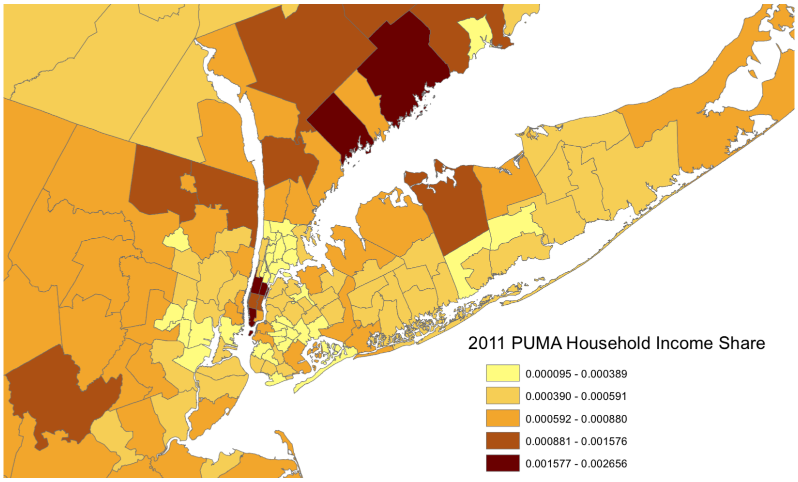

FIGURE 15

Even within Manhattan there is a disparity between PUMAs in income shares, although all are still at the upper end of the income share distribution. As per Figure 8, all these PUMAs have high levels of household income inequality. This corroborates the pattern observed in Table 1—higher inequality areas will tend to have higher shares of the national income.

Figure 16 shows the standard deviation for the annual Household Gini index of PUMAs for years 2005-2011. Lighter areas correspond to higher standard deviations and thus higher rates of inequality fluctuation.

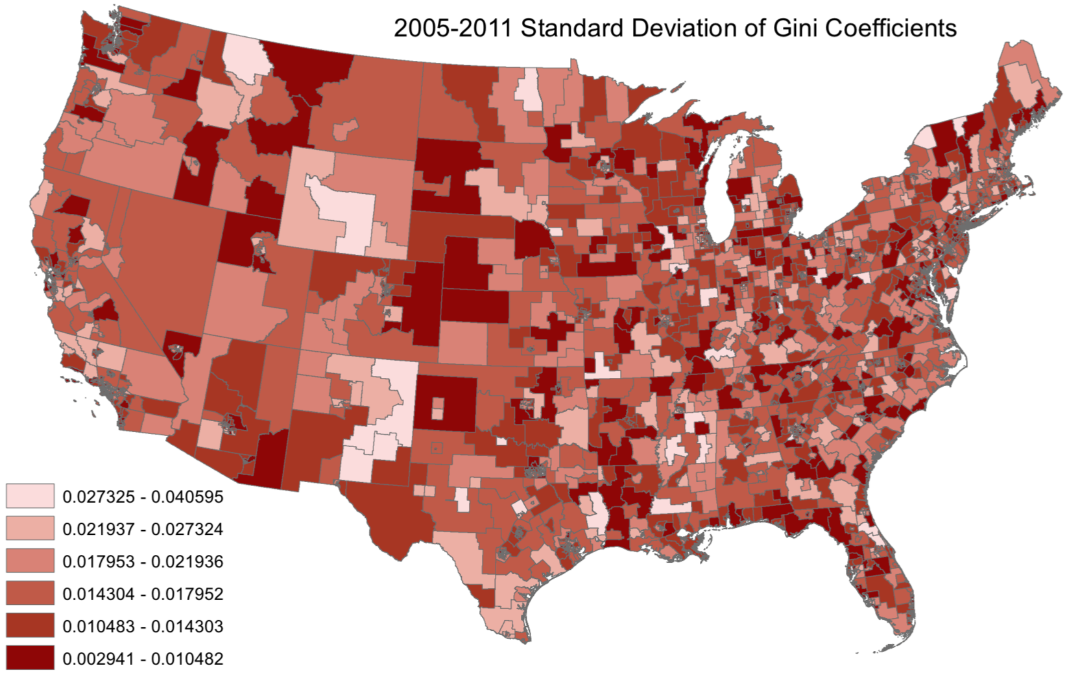

FIGURE 16

This measure is intended to show the locations of high fluctuation in income inequality. While inequality appears to be highly spatially segregated, there is no immediately apparent spatial pattern to the fluctuations of inequality during this time period—the weak cluster of low values along the gulf coast is the lone exception. There are many factors that could lead to different rates of inequality fluctuation. Migration, industry-specific business cycles, and place-specific effects from the Great Recession could also cause a higher-than-normal fluctuation in geographic income. The size of the upper values is notable—the standard deviation of .04 could constitute a fairly large shift in areas with income inequality. This does provide potential target areas for studying the effects of rapidly changing income inequality.

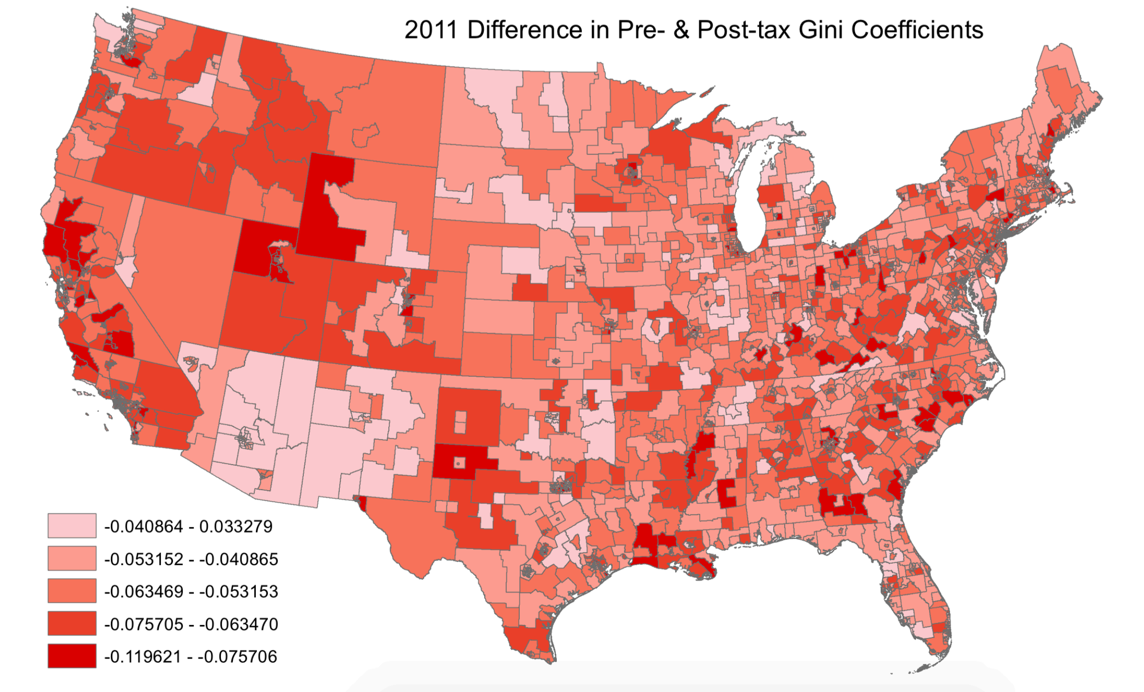

FIGURE 17

Lastly, Figure 17 shows the difference in the PUMA Gini coefficients calculated from the pre-tax and post-tax estimates of household income. The pre-tax estimates are calculated using a summed total of the income input variables after (Samwick 2012)'s transformation. The post-tax Ginis are calculated from this total income variable net the TAXSIM state and federal tax liabilities. The differences shown are the pre-tax values subtracted from the post-tax values.

These can be interpreted in several ways. States such as Arizona or New Mexico, which have near zero or positive absolute Gini changes in their contained PUMAs may have inherently less progressive taxation schemes. However, as these estimates include federal tax liabilities, the populations of these states may also have widespread features on an individual level that make them less likely to qualify for federal transfers. They may also simply have a relative lack of extremely high-income earners, which would prevent any significant reduction in inequality from taxation.

These estimates should still be interpreted with caution. TAXSIM does not calculate local tax liabilities, which may exaggerate the relative inequality of certain areas. For instance, large cities such as Seattle make up for the absence of a state income tax by leveraging higher municipal taxes, which are not accounted for in TAXSIM. Still, if one is specifically interested in examining the distribution of take-home income of households, these estimates likely provide a more accurate picture than measures calculated with raw household Income.

Conclusions

This paper corroborates others that have found a high degree of spatial segregation in American income inequality, both before and after the Great Recession. This inequality is commonly located in denser, urban spaces.

By examining the distributions of Thiel and Gini indices across PUMAs and states, this paper illustrates the presence of a strong influence of geographical aggregation on both the values and the distribution of income inequality indices. Calculating inequality measures at larger levels of geographies is likely to hide high levels of variation in inequality that occur at smaller levels. By attributing a single, often very different level of inequality to a large group of smaller geographies, one essentially paints them all with the same brush. This is problematic especially if inequality does indeed have 'psychosocial effects.' It is not clear whether these operate at small or large areas. If it is the former, then attributing a very different inequality value to the area will likely work to obscure any conclusive analysis. This 'geographic aggregation effect' is in many ways a reiteration of many of the features of the modifiable areal unit problem. Unfortunately, one does not specifically need to be pursuing a spatial analysis in order to fall victim to the MAUP. Because income inequality is a comparative concept, the inclusion or exclusion of individuals based on geography is inherent in any calculation.

There are important policy implications for some of these findings. As the debate over the proper policy response to rising income inequality continues, the spatial element of income inequality has important implications for any policy response. If inequality is to be corrected, it must first be accurately described and located. Many of the opportunity-based solutions must be implemented in the certain areas in order to have any potential effect on income inequality.

There are no established parameters about the geographic level of the 'psychosocial effects' of income inequality, but investigating and establishing it requires accurate, timely data. The accuracy of the income distribution calculations in this paper makes this paper's methods a viable alternative compared to other common sources of inequality estimates.

Identifying areas of high inequality fluctuation is also an important finding of this paper. These are the areas that experienced above-average changes in income inequality during a turbulent economic cycle. By identifying their geographic location, further research into the causes and effects of these high fluctuations in inequality is now possible.

There are several weaknesses of this approach. One is the lack of availability of geospatial PUMS data before 2005. Many analyses attempt to investigate the course of inequality over a much longer time horizon. Current PUMS data is not suitable for such research ventures. Although PUMS data can accurately describe the U.S. income during the Great Recession, it cannot be compared to trends in any other period. Another weakness is the continued redrawing of PUMA boundaries after each decennial census. Although this is a necessary feature of the program, it precludes the use of statistical methods such as place-fixed effects beyond a period of ten years.

PUMA geography is still a relatively new and is largely absent from geospatial data sources that are not affiliated with the US Census. Thus future spatial analyses using the PUMA geography will likely be constrained to using ACS or decennial census data—which is, however, a rich resource. There are some areas in which a PUMA aligns with a specific county or counties. In the absence of non-census sources adopting the PUMA geography, a careful catalog and merge of these 'overlapping' instances would provide a valuable avenue for extending research beyond Census-provided data. For instance, (Samwick 2012) uses a specific geographic overlap in the Alabama PUMAs to combine the PUMS data with information on the county schools.

The repeated significance of urban areas in structuring also inequality invites a broader range of academic disciplines to weigh in on an academic debate largely dominated by social policy and economics. Ideas from economic geography, urban planning, and urban sociology may help inform the investigation of academic inequality. These are disciplines focused on studying urban processes, which clearly play a part in shaping income inequality. While the geographic aggregation effect described in this paper is troubling, the method of use for the ACS PUMS provides a promising resource for inequality research.

References

Bryant, Anthony, Combs, Pierre-Phillippe, and Lafourcade, Miren (2010). "Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? Journal of Urban Economics. Vol 67, Issue 3 pp. 287–302

Feenberg, Daniel and Coutts, Elisabeth (1993). 'An Introduction to the TAXSIM Model', Journal of Policy Analysis and Management, 12 (1), 189–194.

Fotheringham et al (2000). "Quantitative Geography: Perspectives on Spatial Data Analysis." Sage, London.

Gottmann, Jean (1957). "Megalopolis or the Urbanization of the Northeastern Seaboard." Economic Geography. Vol. 33, No. 3 pp. 189-200

Jenkins, Stephen, (2008), Estimation and interpretation of measures of inequality, poverty, and social welfare using Stata, North American Stata Users' Group Meetings 2006, Stata Users Group, http://EconPapers.repec.org/RePEc:boc:asug06:16.

Jenkins, Stephen (2009). "Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution". Review of Income and Wealth, 55: 392–398

Jenkins, Stephen and Van Kerm, Phillipe (2009). 'The measurement of economic inequality', Chapter 3 in W. Salverda, B. Nolan, and Timothy M. Smeeding (ed.s), The Oxford Handbook on Economic Inequality. Oxford: Oxford University Press, 40–67

Openshaw, Tand and Taylor, Peter (1979). "A million or so correlation coefficients: three experiments on the modifiable areal unit problem." In Wrigley N (ed.) Statistical applications in the spatial sciences (Pion, London) 127-144

Perry, Mark (2014). "Sorry Krugman, Piketty and Stiglitz: Income inequality for individual Americans has been flat for more than 50 years." American Enterprise Institute. https://www.aei.org/publication/sorry-krugman-piketty-and-stiglitz-income-inequality-for-individual-americans-has-been-flat-for-more-than-50-years/ Accessed 10 April 2015

Peters, David (2012). “Understanding micro and meso scale income inequality in the Midwestern United States, 1979–2009.” Rural Sociology 77, 171–202.

Peters, David (2013). “American Income Inequality Across Economic and Geographic Space, 1970–2010” Social Science Research 42: 1490–1504

Pickett Kate E., and Wilkinson Richard G (2015). “Income Inequality and Health: A Causal Review” Social Science & Medicine 128: 316-326

Piketty, Thomas (2014). “Capital in the Twenty-First Century.” Cambridge, MA: Harvard University Press

Posey Kirby G., Welniak Edward, and Nelson, Charles (2003). “Income in the American Community Survey: Comparisons to Census 2000.” U.S. Census Bureau. Presented to American Statistical Association Meetings August 2002 San Francisco, California

Samwick, Andrew A., (2013). "Donating the Voucher: An Alternative Tax Treatment of Private School Enrollment," Tax Policy and the Economy, University of Chicago Press, vol. 27(1), pages 125 - 160.

Thompson, Jeffrey and Smeeding, Timothy (2013). “Country case study – USA.” In Jenkins, S.P., A. Brandolini, J. Micklewright and B. Nolan, eds. (2013), “The Great Recession and the Distribution of Household Income”, Oxford University Press, Oxford.

University of Wisconsin Population Health Institute (2015). “County Health Rankings Key Findings Report.” Madison, WI

U.S. Census Bureau (2000). “Guidelines for the Delineation of 5-Percent and 1-Percent Public Use Microdata Areas (PUMAs)”. U.S. Government Printing Office, Washington, DC.

U.S. Census Bureau (2003). “Census 2000 Summary File 4 Technical Documentation” U.S. Government Printing Office, Washington, DC.

U.S. Census Bureau (2007). “American Community Survey and Puerto Rico Community Survey 2007 Subject Definitions” U.S. Government Printing Office, Washington, DC.

U.S. Census Bureau (2009). "A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know." U.S. Government Printing Office, Washington, DC.

U.S. Census Bureau (2011). "American Community Survey and Puerto Rico Community Survey 2011 Subject Definitions" U.S. Government Printing Office, Washington, DC.

Weinberg, Daniel H. (2011). "U.S. Neighborhood Income Inequality in the 2005–2009 Period". American Community Survey Reports, Issued October 2011. US Census Bureau.

Wheeler, C., La Jeunesse, E., (2008). "Trends in Neighborhood Income Inequality in the U.S.: 1980–2000." Journal of Regional Science 48, 879–891

Wilkinson, Richard. G., & Pickett, Kate. (2010). "The spirit level: Why greater equality makes societies stronger." New York: Bloomsbury Press.