Re-examining Regional Income Convergence: A Distributional Approach*

Kevin Rinz and John Voorheis

This version: February 2, 2023
Current version available here

Abstract

We re-examine recent trends in regional income convergence, considering the full distribution of income rather than focusing on the mean. Measuring similarity by comparing each percentile of state distributions to the corresponding percentile of the national distribution, we find that state incomes have become less similar (i.e. they have diverged) within the top 20 percent of the income distribution since 1969. The top percentile alone accounts for more than half of aggregate divergence across states over this period by our measure, and the top five percentiles combine to account for 93 percent. Divergence in top incomes across states appears to be driven largely by changes in top incomes among White people, while top incomes among Black people have experienced relatively little divergence.

*Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release. Disclosure Review Board release authorization numbers CBDRB-FY23-051 and CBDRB-FY23-0136. Rinz: U.S. Census Bureau and Federal Reserve Bank of Minneapolis, kevin.rinz@census.gov; Voorheis: U.S. Census Bureau, john.l.voorheis@census.gov.

1 Introduction

Consistent with predictions of canonical growth models, prior research has found that incomes have converged across American states since at least the late 19th century (Barro and Sala-i Martin, 1991). This convergence has been realized in the form of reduced dispersion in mean incomes across states (so-called σ convergence), as well as faster growth in income in states with lower initial income levels (β-convergence), but the rate of convergence has slowed in recent decades, perhaps due to increased housing costs and reduced low-skill migration (Ganong and Shoag, 2017). However, prior work on regional convergence has focused largely on measures of average income, which may not fully capture differences across states throughout the income distribution, especially during periods of high or rising inequality.[1] This is particularly salient given the increases in inequality observed over the same period that saw the rate of convergence decline. As shown in Figure 1, the average wage and salary income of the top one percent of tax units increased from $281,000 in 1969 to $796,000 in 2019, nearly tripling in real terms.[2] Over the same period, the bottom 70 percent of the income distribution saw little change, with some parts of the income distribution experiencing real declines over those 50 years. Moreover, Giannone (2022) finds that college educated workers' wages diverged across cities between 1980 and 2010 (rather than continuing to converge as they had between 1940 and 1980, and as the wages of workers without college degrees did in both periods), and this has driven the overall slowdown in convergence since 1980. Would re-examining trends in regional income convergence with an eye toward assessing distributional differences more directly enhance our understanding of the recent history of this phenomenon? More specifically, would measuring the similarity of state income distributions percentile by percentile over time reveal important patterns not captured by analysis of averages? The answer is yes: top incomes have played a much more significant role in the evolution of regional convergence trends than previously appreciated.

We address this question using the universe of tax returns for select years from 1969 to 2019. The high-resolution picture of the income distribution provided by these datasets allows us to build on prior work and construct measures that compare states' income distributions percentile by percentile rather than relying on summary measures. We estimate cumulative distribution functions (CDF) for income in each state and assess convergence by evaluating differences between each state's CDF and the national distribution of income.

[1] In addition to works cited above, see, for example, Rey and Montouri (1999); Tomljanovich and Vogelsang (2002); Hammond and Thompson (2006).

[2] Amounts are adjusted to 2019 dollars using the CPI-U price index.

Comparisons in this style can also be applied to subgroups of state populations, including those defined by demographic characteristics.[3] Alongside this paper, we are releasing detailed measures of state and national income distributions for 1998–2019, including estimates by gender and by race/ethnicity.[4] These measures can be used to conduct a wide variety of analyses, as we illustrate below, including convergence analysis using the approach described here.

Using tax data to analyze the relationship between mean income levels and growth does not meaningfully alter the recent picture of relatively slow regional convergence. For example, the relationship between mean income level in 1998 and income growth over the subsequent 21 years is small in magnitude and not statistically significant using both adjusted gross income and a publicly available measure like per capita personal income. However, extending this style of analysis to percentiles of the income distribution does provide new insight into recent patterns of convergence and divergence. In general, over the same period that saw little convergence in mean incomes, state income distributions converged more quickly at percentiles in the bottom four quintiles of the distribution while generally diverging in the top quintile.

Analyzing convergence using the average absolute deviation of income within percentiles of state income distributions from income within national percentiles further confirms that considering average convergence masks important heterogeneity in the experience of convergence across the income distribution. State distributions of wage and salary income became more similar between 1969 and 1979 before they began to diverge. That divergence, however, is driven by the top of the income distribution. States remain more similar to each other within the bottom 80 percent of the income distribution than they were in 1969, while they have become less similar to each other within the top 20 percent, with that dissimilarity increasing the further up the income distribution we look. The top percentile alone accounts for more than half of the average divergence across states by this measure since 1969, and the top five percentiles combine to account for 93 percent. Furthermore, analysis conducted within groups defined by race suggests that divergence in top incomes across states is driven largely by the incomes of White people. Top incomes among Black people have experienced relatively little divergence. These differential changes in similarity across states at

[3] While one could, of course, use the same underlying microdata to directly compare means of outcomes of interest across groups, we hope that the minimal structure we apply to comparisons across groups can make them more comprehensible, as well as provide value in settings in which the relative incomes of demographic groups and their members are of interest in and of themselves. It is also possible to do similar exercises using survey data from the American Community Survey or the long-form Decennial Census (for instance Manduca 2019) – our approach provides advantages over this approach, both in fidelity at the top of the income distribution, and in the ability to consistently explore differences within even small sociodemographic groups across states.

[4] Data are released starting in 1998 because this is the beginning of the period for which the underlying microdata are available annually.

different points in the income distribution and across demographic groups should help provide guidance to future work looking to identify causes of changes in regional convergence.

The rest of this paper is organized as follows. Section 2 discusses the data. Section 3 presents evidence on the relationship between income levels and subsequent income growth across the distribution (β-convergence). Section 4 defines the measure we use to analyze convergence across the full distribution of income, as well as related measures that can be used to describe differences across groups and places more broadly, and illustrates a few additional uses of those related measures. Section 5 examines regional income convergence across the full distribution. Section 6 concludes.

2 Data

We use the U.S. Census Bureau's data linkage infrastructure to combine a variety of data sources to create an analysis dataset which provides detailed demographic characteristics, income from tax data, and residential history information from administrative records.[5]

2.1 List of people from the Social Security Administration Numident file

The Social Security Administration's Numident file records all transactions against all Social Security Numbers (SSN). Through the Census Bureau, we have access to an extract from this file that consists of one record per SSN and provides information such as date of birth, gender, place of birth, and date of death, if applicable. The Census-assigned Protected Identification Key (PIK), an anonymized person identifier used for linking data across files, is also on this file. The Numident provides the population to which we attach income, demographic, and location data from other sources. We use dates of birth and death to focus on only people who were alive in each year we consider.

2.2 Income measures from tax data

We obtain income measures from IRS Form 1040. Tax-based income measures have the advantage of capturing the top tail of the income distribution more accurately than survey-based measures, and the availability of the universe of tax records will allow the extension of the approach described here to finer levels of geography than would be possible with sample surveys. Extracts available through Census provide select fields

[5]For more on data linkage at the Census Bureau, see Wagner and Layne (2014).

from these forms for the universe of records. The 1040s, which are available annually since 1998 and in select years prior to that back to 1969, provide tax-unit level measures of adjusted gross income (AGI) and wage and salary income. In some parts of this analysis, we focus on wage and salary income because the reporting of this type of income was less likely than a broader measure like AGI to be influenced by changes in tax law such as the Tax Reform Act of 1986. Using administrative tax data is important for understanding differences across state distributions, particularly at the top of the income distribution. Survey-based estimates of the income distribution will systematically underestimate top incomes, both because of disclosure avoidance (topcoding) and survey misreporting (Burkhauser et al., 2011; Bee and Mitchell, 2017).[6]

2.3 Demographics

For age and gender, we use data from the Numident, given its availability for essentially all records. For race and ethnicity, we use data from the Census Bureau's Best Race and Ethnicity Administrative Records Composite File, which combines information on race and ethnicity from Census Bureau data collections like the decennial census and American Community Survey with data from various administrative and commercial sources into a single composite file. Except where otherwise noted, we combine race and ethnicity into 6 mutually exclusive groups: Hispanic of any race, Non-Hispanic White alone, Non-Hispanic Black alone, Non-Hispanic American Indian and Alaska Native alone, Non-Hispanic Asian or Native Hawaiian/Pacific Islander alone, and Non-Hispanic Some other race or two or more races.[7]

2.4 Location

We obtain data on place of residence from IRS Form 1040. Each return is assigned to the state from which it was filed.

3 β-convergence across the income distribution

We begin our analysis of regional convergence by considering the relationship between income levels and subsequent income growth (so-called β-convergence) across the income distribution. To provide an initial

[6] Tax data do have the limitation of not capturing the incomes of people who do not file tax returns. If rates of non-filing differ across places, groups, or time, comparisons of income distributions using tax data could partially reflect those differences in whose income is reflected.

[7] We will refer to these groups as "Hispanic", "White", "Black", "AIAN", "Asian" and "Other/Multiracial" for brevity. In Section 5.1, we analyze White and Black groups that are constructed based on race alone due to changing availability of data on Hispanic ethnicity over time.

point of comparison with other analyses, Figure 2 plots income levels in 1950 against income growth over the subsequent 21 years using per capita personal income (PCPI), as well as that same relationship for the period beginning in 1998 using both PCPI and AGI.[8] Comparing the two PCPI plots illustrates the degree to which β-convergence has slowed in recent decades using a publicly available measure of mean income. Comparing the 1998 plots using PCPI and AGI illustrates that the two data sources paint a similar picture of β-convergence over the same time period.[9]

Figure 3 depicts β-convergence between 1998 and 2019 across the AGI distribution. Aside from volatility at the bottom of the distribution, the relationship between initial income level and subsequent growth rates is consistently negative and fairly stable below roughly the 70th percentile. The coefficient then begins to diminish in magnitude before turning positive in the 84th percentile. Throughout a large majority of the income distribution, states experienced modest β-convergence: those with lower values of a given percentile experienced somewhat faster growth. At the top of the distribution, however, higher-income states are consistently growing faster, implying that states are becoming less similar in this portion of the distribution.[10]

4 Measuring distributional differences more directly

The relationship between initial income levels and subsequent growth provides only an indirect view into whether state income distributions are becoming more or less similar over time, even when it is considered across many percentiles. We use tax data to construct measures that compare income distributions across places directly, allowing for a more intuitive assessment of regional convergence.

4.1 Assessing convergence using absolute differences

We use absolute differences to measure dissimilarity between state income distributions and the national distribution at each percentile and then use changes in this dissimilarity measure over time to assess convergence. Let $F_N(y)$ be the cumulative national income distribution for an income concept y, and $F_n(y)$ be

[8] Personal income includes income received by persons as compensation for employment or from sources like business ownership, rental property, social security, other government benefits, interest, and dividends. Adjusted gross income includes income from all sources but is reduced by the amount of various adjustments made for tax purposes (e.g. the amount of student loan interest or alimony paid). Here and throughout this paper, we condition our analysis of tax data on returns reporting positive values of the income measure in question.

[9] Among the available tax-based measures, AGI is most comparable to PCPI in that both include earnings from labor as well as other sources. We do not have access to tax data for the period beginning in 1950, and the later periods for which we do have tax data do not exhibit faster β-convergence.

[10] Performing this exercise using wage and salary income instead of AGI produces qualitatively similar results.

the nth region's cumulative income distribution. We denote the quantile function as $q_n(p)$ for the regional distribution, and $q_N(p)$ for the national distribution. For each region n, we then calculate the absolute value of the difference between the value of the p^{th} percentile in region n and that same percentile nationally:

$$N_n(p) = |q_n(p) - q_N(p)|, \forall p \in [0,1] \tag{1}$$

Integrating this measure over the full distribution then gives the average absolute difference between the regional and national distributions.

$$AAD_n = \int_0^1 N_n(p) f_n(p) dp \tag{2}$$

We then use the values of AAD_n for all states to assess similarity between state income distributions. The average of AAD_n across states can be thought of as a dissimilarity measure: the higher the value of this average, the less similar to each other are state income distributions. This maps onto discussion of income convergence or divergence quite naturally. If state income distributions converge, becoming more similar to the national distribution (and by extension each other), this value falls; if they diverge, it rises. Assessing regional convergence in this way is analogous to evaluating the national income distribution as an estimator of state income distributions based on its mean absolute error.

4.2 More general measures of distributional differences

We can also use a similar approach to characterize more general differences in income distributions across places or groups. These measures can be found in appendix files accompanying this paper. See Appendix B for details.[11]

4.2.1 Regional Income Divergence Curves

We define two related curves:

Definition 1 (Relative Regional Income Divergence Curve)**.** $R(p) = F_N(q_n(p)) - p, \forall p \in [0,1]$

Definition 2 (Absolute Regional Income Divergence Curve)**.** $A(p) = q_N(p) - q_n(p), \forall p \in [0,1]$

The Relative Regional Income Divergence Curve (RRIDC) traces out the percentile rank difference between the regional income distribution and the national income distribution, while the Absolute Regional

[11] Data files are available here.

Income Divergence Curve (ARIDC) traces out the dollar value difference between the regional income distribution and the national income distribution. These can be visualized as aggregating the vertical (RRIDC) or horizontal (ARIDC) differences between any two CDFs.

4.2.2 Regional Income Divergence Indices

The RRIDC and ARIDC are useful tools for visualizing differences between regional and national income distributions, and could, for instance, be used in stochastic-dominance type analyses; however this may be less useful for social evaluation, as this may not yield unambiguous rankings across distributions. We can, however, aggregate these curves to indices which can be used as social evaluation functions.

First, consider two simple aggregations of these income divergence indices which calculate the unweighted mean of the income divergence across the regional income distribution:

$$D_R(p) = \int_0^1 R(p)f(p)dp$$

$$A_R(p) = \int_0^1 A(p)f(p)dp$$

On their own, these aggregations are not particularly useful as social evaluation functions, for reasons paralleling the arguments in (Van Kerm, 2009). However, we can produce aggregates with nice normative features by applying an appropriate weighting scheme. Similar in structure to the income mobility profiles and associated indices, we define two (Yaari, 1987)-style indices:

Definition 3 (Divergence Indices)**.**

$$RDI(p) = \int_0^1 w(p)R(p)dp$$

$$ADI(p) = \int_0^1 w(p)A(p)dp$$

With the weighting function $w(p)$, such that $\int w(p)dp = 1$. One logical choice for a weighting scheme would put higher weight on lower (regional) incomes, consistent with a Rawlsian social evaluation function. One such weighting scheme would be to use Gini weights such that $w(p) = \nu(1-p)^{\nu-1}$.

These divergence indices have important relationships with $D_R(p)$ and $A_R(p)$ which can be used to draw normative conclusions. First, $RDI(p)$ and $ADI(p)$ can themselves be interpreted as the equally distributed equivalent of the $D_R(p)$ and $A_R(p)$ respectively. That is to say, it is the amount of divergence which, if

it were constant across the income distribution would be just the same as the actual pattern of divergence. Similarly, the difference between our indices and the unweighted average aggregations can be interpreted as the total welfare loss (or gain) of the observed pattern of divergence relative to no regional divergence (i.e. where each region's income distribution mirrors the national distribution), as in Van Kerm (2009).

4.3 Example uses of new distributional measures

Information on local income distributions, including distributions within various demographic groups, has a wide wide variety of potential uses. While providing an exhaustive catalog or detailed analysis of the possibilities is beyond the scope of this paper, we provide a few examples to illustrate how the measures defined above could be broadly useful for making comparisons across demographics, place, and time.[12]

4.3.1 Relative incomes by race and ethnicity in high- and low-income places

Some state income distributions are situated further in the right tail of the national income distribution than others. To what extent is this fact driven by the income distributions of demographically defined subgroups of their populations? In a high-income state like Connecticut, are both the White and Black income distributions to the right of the overall national distribution? Are they each to the right of the corresponding distributions in a low-income state like Mississippi?

Figure 4 plots relative regional income divergence curves by race and ethnicity for Connecticut (the state with the highest per capita personal income) and Mississippi (the state with the lowest) for 2019. Curves are constructed using AGI. Despite the state's high average income, the distributions of income for Connecticut's Black and Hispanic populations were substantially to the left of the overall national distribution: Black and Hispanic residents of Connecticut who were at the 70th percentile of their own group's income distribution in the state were only just above the national median income. White and Asian/Pacific Islander income distributions in Connecticut sat substantially to the right of the national distribution. Residents around the medians of their own groups' distributions were closer to the 65th to 70th percentiles of the overall national distribution. In Mississippi, the relative standing of these groups is similar, but they are situated lower in the national distribution. White and Asian incomes are roughly in line with the national distribution while Black and Hispanic incomes are well below. The 80th percentile Black Mississippian is below the national

[12]Note that these analyses cannot necessarily be reproduced using the data released alongside this paper and are intended to demonstrate the utility of this mode of analysis more generally.

median.

These two states suggest that being in a high-income state may advantage lower-income groups such as Black and Hispanic residents relative to lower-income residents of low-income states, but not necessarily relative to higher-income residents of low-income states. Figure 5 confirms that Connecticut and Mississippi are not anomalous and that this fact has held over time. The figure plots the range of RRIDCs observed in the ten highest-income states and the ten lowest-income states for White and Black residents in 1998 and 2019. In each case, there is a clear ordering of groups: White residents of high-income states, White residents of low-income states, Black residents of high-income states, and Black residents of low-income states, though there is substantial overlap between the Black income distributions in high- and low-income states. Moreover, White income distributions tend to be in line with or to the right of the national income distribution, while Black income distributions tend to be in line with or to the left of the national distribution. In short, even in the lowest-income states, White incomes are still in line with the national distribution, while even in the highest-income states, Black incomes are only in line with the national distribution.

4.3.2 Two decades of change in the industrial Midwest

Employment in motor vehicle and parts manufacturing peaked around 1.3 million in the late 1990s and into 2000, and about half of these jobs were located in Indiana, Michigan, and Ohio. More broadly, about 3.1 million people were employed across the manufacturing sector in these states in 1998, representing more than 20 percent of total employment. Each of those states was a relatively good place to be for workers that year, and for men in particular, about 2.2 million of whom held manufacturing jobs. Adjusted gross income distributions were to the right of the overall national distribution for the bottom 60 percent of Ohio men, the bottom 75 percent of Indiana men, and all Michigan men, as shown in Figure 6.

By 2019, however, after changes in trade policy increased competition from foreign manufacturers (Autor et al., 2013) and broader economies only partially recovered from two recessions that hit the region particularly intensely (Hershbein and Stuart, 2022), circumstances had changed. While men in the broad middle class and below had been advantaged in these states relative to the national income distribution in 1998, the income ranges in which that was true in 2019 were much more limited. Only the bottom 40 percent or so of men in Indiana and the bottom 47 percent of Ohio men were situated higher in the overall national distribution than in their own group's distribution. The most dramatic shift took place in Michigan, where essentially the full distribution of men's incomes sat to the right of the national distribution in 1998.

By 2019, Michigan men's incomes were generally only in line with or below overall incomes nationally.

4.3.3 Longer-term not-so-shared prosperity

Income gaps between White and Black Americans are longstanding and persistent. Panel (a) of Figure 7 illustrates that among prime-age people in 1969, the earliest year in our data, the average Black tax-filer in nearly every state had lower wage and salary income than the average American (Alaska was the only exception), with the gap being notably larger within the states of the American South, which were and are home to a large share of Black Americans.[13] Panel (b) shows that, in the same year, White tax-filers had more varied experiences across states, with some states above and others below the national average. By 2019, overall incomes had grown substantially, but so had inequality, and Black Americans in more places were left further behind, as shown in panel (c), while Whites in coastal states like California in the west and Connecticut and New Jersey in the east out-earned the average American by much more than they had decades earlier, as shown in panel(d).

The degree to which states tended to have higher incomes for both their White and Black residents also changed over time. Figure 8 plots the Black absolute divergence index on the y-axis against the White absolute divergence index on the x-axis, again focusing on prime-age people. In 1969, depicted in panel (a), Black incomes were clearly lower than White incomes in all states except Vermont. In a state where White residents had average income in line with the national average, Black residents had average income nearly $19,000 below it (in 2019 dollars). States with higher White incomes tended to also have higher Black incomes, but the slope of the regression line was only about 0.5. By 2019, however, the gap between White and Black incomes had increased (the regression line's intercept term had increased to about $30,000), and White incomes were less correlated with Black incomes (the slope term had decreased to about 0.3).[14] More than 50 years after the civil rights legislation of the 1960s attempted to reduce de jure segregation, the economic circumstances of White and Black Americans seem less connected by these measures rather than more so.

[13] Here we focus on wage and salary income because the period covered by this analysis includes the Tax Reform Act of 1986, which likely affected AGI reporting by shifting corporate income into pass-through entities. We also focus on prime-age workers to limit the influence of the aging of the population on these comparisons.

[14] Constructing these measures with Gini weights that increase the weight placed on lower-income people ($\nu = 5$) shows less change between 1969 and 2019. See Figures A1 and A2 in Appendix A.

5 Convergence Across the Full Distribution

We now use the average absolute difference measure described in Equation 2 to evaluate whether/to what extent state income distributions have become more similar over time. Figure 9 plots key moments of the distribution of AAD_n, which is constructed using wage and salary income and estimated for each state using equal weight on all percentiles.[15] Three key facts are evident. First, the average of AAD_n across states fell slightly between 1969 and 1979 but has risen since then, suggesting that state income distributions have become less similar to each other by this measure over the last 40 years. This overall pattern is consistent with results from some measures used in prior work over this time period, providing some validation of our measure.[16] Second, all percentile values depicted in the figure's box plots are greater in 2019 than they were in 1969, suggesting that the increase in the mean of AAD_n across states is not driven by a few states diverging from the others dramatically while most become more similar. Finally, the difference between the 10th and 90th percentiles is also greater in 2019 than it was in 1969, suggesting some potential role for increased inequality across states in the divergence observed over this period.

Decomposing the average of AAD_n across states into contributions from various regions of the income distribution also suggests a role for inequality in its evolution. Figure 10 shows that income divergence across states has been driven by the top of the income distribution. Panel (a) shows that the contribution of the top one percent of state income distributions to their overall dissimilarity has increased substantially and steadily since 1979, with the contribution of the 95th through 99th percentiles also increasing meaningfully. In contrast, the bottom 80 percentiles of state income distributions were responsible for the overall convergence across states between 1969 and 1979, and states remain more similar in this region of the income distribution than they were in 1969. Panel (b) makes the contributions of these income groups to overall income divergence over this period clear. The top percentile was responsible for 57 percent of overall income divergence across states between 1969 and 2019 under this measure. Including the 95th through 99th percentiles, the top five percent of the income distribution accounted for 93 percent of income divergence.[17]

Revisiting the β-convergence results reported in Figure 3 in light of this analysis illustrates an important

[15] Despite the conceptual break in series introduced by the Tax Reform Act of 1986, performing this exercise using AGI yields qualitatively similar results.

[16] The dispersion of personal income across states, for example, declined between 1969 and 1979. See Figure 4 in Barro and Sala-i Martin (1991).

[17] Repeating this analysis using public-use decennial census and American Community Survey data on wage and salary income from IPUMS (Ruggles et al., 2022) also shows divergence over this period, but it is much more broadly experienced throughout the distribution. The public-use version of this analysis does not show a similarly key role in income divergence for the top percentile. See Figures A3 and A4.

benefit of considering regional income convergence more comprehensively across the income distribution. Figure 3 might have suggested some notable role for the top of the income distribution in recent trends in regional income convergence. It would not necessarily have pointed, however, to the central role of the top percentile in particular. For example, the top percentile experienced essentially no β-convergence (or divergence) between 1998 and 2019 (as shown in Figure 3), a period during which its contribution to divergence by our measure was substantial. The approach developed here more naturally reflects changes in the similarity of state income distributions over time and reveals this key fact about the importance of the very top of the income distribution to understanding trends in regional income divergence over the last several decades.

Our average absolute difference measure of distributional similarity enables us to identify the trends discussed here and build on the understanding of regional income convergence established by prior research, but it is not the only measure one might use to assess distributional similarity in this spirit. For example, a measure capturing something analogous to the commonly used mean squared error concept could also average differences between state and national income distributions across percentiles, but would penalize large differences more severely than our measure does. Because the largest differences between state and national distributions tend to occur at top percentiles, using a measure like this would show more substantial income divergence and a larger role for the top percentile in driving it than our measure does (see Appendix Figure A5). Our approach is, relative to this alternative possibility, conservative with respect to attributing changes in distributional similarity to the top percentile.

5.1 Distributional Convergence by Race

Since the distributions of income for White and Black households differ substantially, particularly in the top tail, one might suspect that the patterns of divergence over time for the overall income distribution may not fully reflect the experience of each of those groups. We can investigate convergence patterns across the distribution by race by comparing state-race income distributions to national distributions using an approach analogous to the one described in Section 4.1.[18]

[18] Unlike the rest of this paper, in this section we use groups defined by race alone, not race and ethnicity. Hispanic individuals who are also White or Black are included in this section's analysis. We took this approach because data on Hispanic ethnicity are much more likely to be missing for earlier years of data, but data on race are still widely available in those years. Using groups defined by race alone for this analysis improves consistency within groups over time. Moreover, since we are not using a Hispanic group for this analysis, the combined White and Black populations constitute an overwhelming majority of the country's population, especially at the beginning of the period considered, so we focus on only those two groups.

We first construct AAD_n measures by race by comparing each state's White and Black wage and salary income distributions to the overall national income distribution. Box plots of these measures across all states are shown in panels (a) and (b) of Figure 11. We first note that differences from the national income distribution throughout the distribution of AAD_n are larger for Black tax filers than for White tax filers. This reflects the fact that 1) a sizable majority of tax filers are White, so the White and overall distributions of income are mechanically similar to some extent, and 2) Black tax filers have lower incomes through the distribution.

Second, panels (a) and (b) suggest that incomes have diverged across states for both White and Black tax filers since 1969. The divergence is notably larger in absolute terms for Black tax filers, though as a proportion of the initial level of dissimilarity, the divergence is larger for White tax filers. However, the higher average levels of difference from the national distribution for Black tax filers discussed above can make changes over time in this measure difficult to interpret. Consider two examples. First, suppose that 25 states with higher-income Black populations became uniformly higher income by some amount across the income distribution. The Black income distributions in these states would become less similar to those in the other states that did not experience this increase but more similar to the national income distribution, since Black incomes are lower than overall incomes throughout the distribution across states. This would lead the Black *AAD* measure to fall even though Black income distributions are clearly less similar to each other across states.

Second, suppose incomes at the top of the White income distribution increase by the same proportion in all states while Black income distributions are unchanged. This will make White income distributions less similar to each other across states, but it will also raise incomes at the top of the overall income distribution, making Black income distributions less similar to it. Black income distributions will appear to have become less similar according to the *AAD* measure even though they have not changed. In order for the approach described in Section 4.1 to make sense when applied to understanding race-specific income divergence, state distributions need to be compared to something that approximates the central tendency of state income distributions (e.g. the national distribution for the same race) rather than something that all states differ from systematically.

To address this we compare state-race income distributions to the national distributions of income for the same race in panels (c) and (d) of Figure 11. Under this approach, White incomes have diverged across states in much the same way they do in comparison to the overall national income distribution in panel (a).

This makes sense, given that White tax filers make up a large share of all tax filers. State Black income distributions, however, are about as similar to the national Black income distribution in 2019 as they were in 1969, and the degree of similarity experienced minimal fluctuation between those years. This suggests that income divergence since 1969 has been driven almost entirely by changes at the top of White income distributions across states.

Figure 12 strengthens this suggestion. The top percentiles of White income distributions have made substantial contributions to state divergence from the national White income distribution, with other percentiles above the median also contributing to divergence. The top percentiles of Black income distributions also make the largest contributions to divergence from the national Black income distribution, but the magnitude of that contribution is only about as large as the 81-90 percentiles of the White distribution combined. Moreover, the bottom 80 percent of Black income distributions have converged across states since 1969, largely offsetting divergence at higher percentiles.[19]

6 Conclusion

Re-examining regional income convergence using a distributional approach reveals that prior analyses focused on mean-based β-convergence have missed important facts about how state income distributions have changed with respect to each other – and the country as a whole – over time. States have become less similar to each other within roughly the top 20 percent of the income distribution over the last four decades. This change has been most pronounced within the top percentile of the distribution, which accounts for more than half of the divergence observed across states using our measure during the period covered by our analysis.

At the same time, our distributional approach confirms that the trends laid out in the prior literature do hold for the rest of the distribution. States became more similar to each other within the bottom 80 percent of the income distribution through the 1970s, but progress on this front has stalled more recently.

Given the many factors that can lead state income distributions to converge or diverge, the more nuanced picture of recent trends produced by our approach should help inform theory on this subject. If land use restrictions are slowing regional convergence by reducing migration among low-earning workers as in Ganong

[19] One might also be interested in examining changes in income distribution similarity across states within groups defined by gender. However, because the income measure used here is defined at the tax unit level and the overwhelming majority of jointly filed tax returns represent (at least) a man and a woman, the current setting is not ideal for identifying gender-specific contributions to changes in convergence trends. Such an analysis could be undertaken using individual-level tax forms like W-2s, but those data are only available to us beginning in 2005.

and Shoag (2017), for example, is that consistent with divergence at the top of state income distributions and slow convergence/stasis further down the distribution? Where in the distribution might changes in the return to college, capital mobility, or agglomeration effects be important for explaining observed patterns of convergence or divergence? Future research should explore these questions further.

References

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–2168.

Barro, Robert J. and Xavier Sala i Martin, "Convergence," *Brookings Papers on Economic Activity*, 1991, *1991* (1), 107–182.

Bee, Adam and Joshua Mitchell, "Do Older Americans Have More Income Than We Think?," *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 2017, *110*, 1–85.

Burkhauser, Richard, Shuaizhang Feng, Stephen P. Jenkins Jenkins, and Jeff Larrimore, "Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring," *Journal of Economic Inequality*, 2011, *9*, 393—-415.

Ganong, Peter and Daniel Shoag, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90.

Giannone, Elisa, "Skilled-biased technical change and regional convergence," *Working paper*, 2022.

Hammond, George W. and Eric Thompson, "Convergence and Mobility: Personal Income Trends in U.S. Metropolitan and Nonmetropolitan Regions," *International Regional Science Review*, 2006, *29*, 35 – 63.

Hershbein, Brad and Bryan A. Stuart, "The Evolution of Local Labor Markets After Recessions," March 2022. Working paper.

Manduca, Robert A, "The Contribution of National Income Inequality to Regional Economic Divergence," *Social Forces*, 03 2019, *98* (2), 622–648.

Rey, Sergio Joseph and Brett D. Montouri, "US Regional Income Convergence: A Spatial Econometric Perspective," *Regional Studies*, 1999, *33*, 143–156.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek, "IPUMS USA: Version 12.0," 2022.

Tomljanovich, Marc and Timothy J. Vogelsang, "Are U.S. regions converging? Using new econometric methods to examine old issues," *Empirical Economics*, 2002, *27*, 49–62.

Van Kerm, Philippe, "Income mobility profiles," *Economics Letters*, 2009, *102* (2), 93–95.

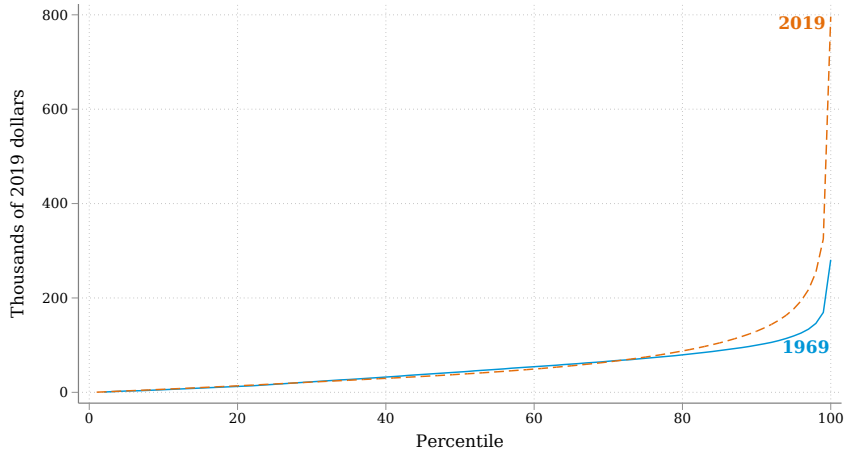
Wagner, Deborah and Mary Layne, "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," July 2014. CARRA Working Paper No. 2014-01.

Yaari, Menahem E., "The Dual Theory of Choice under Risk," *Econometrica*, January 1987.

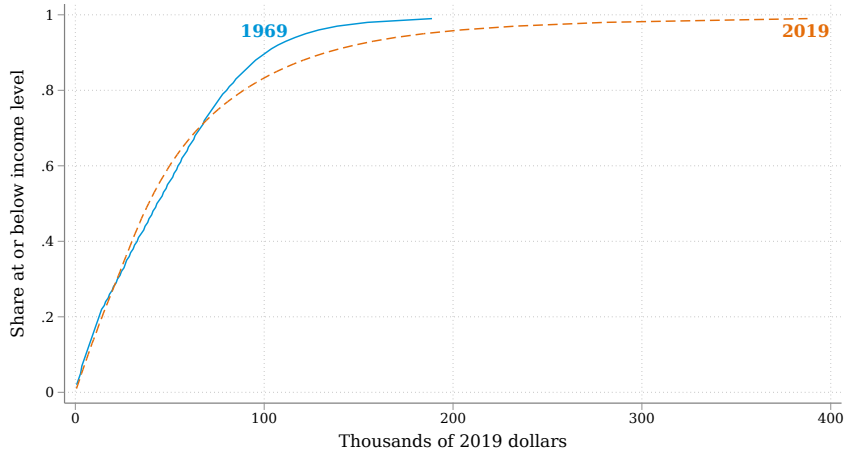
Figures

Figure 1: Distribution of wage and salary income, 1969 and 2019

(a) Average wage and salary income by percentile

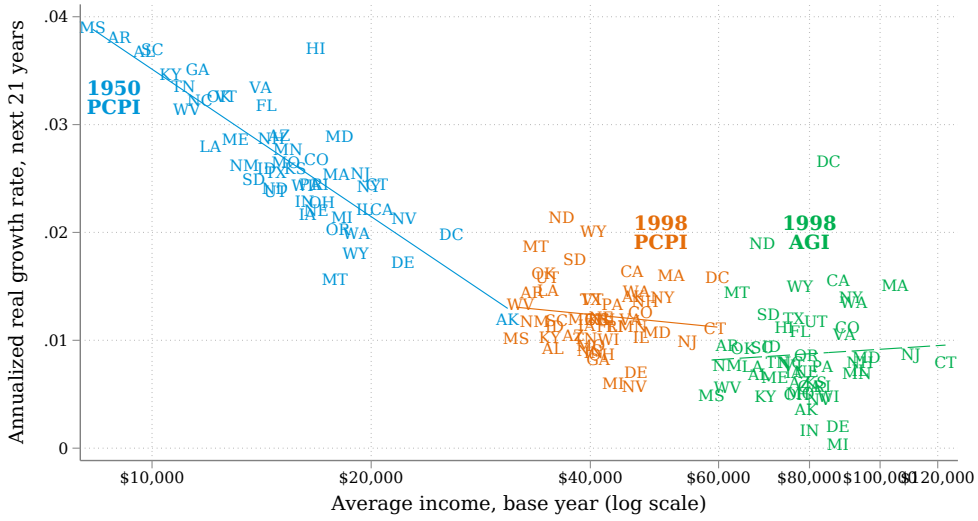


(b) Cumulative distribution function



Source: IRS Form 1040 and authors' calculations.
Note: Values are in 2019 dollars, adjusted for inflation using the CPI-U. Analysis is at the tax unit level. Release authorization CBDRB-FY23-051.

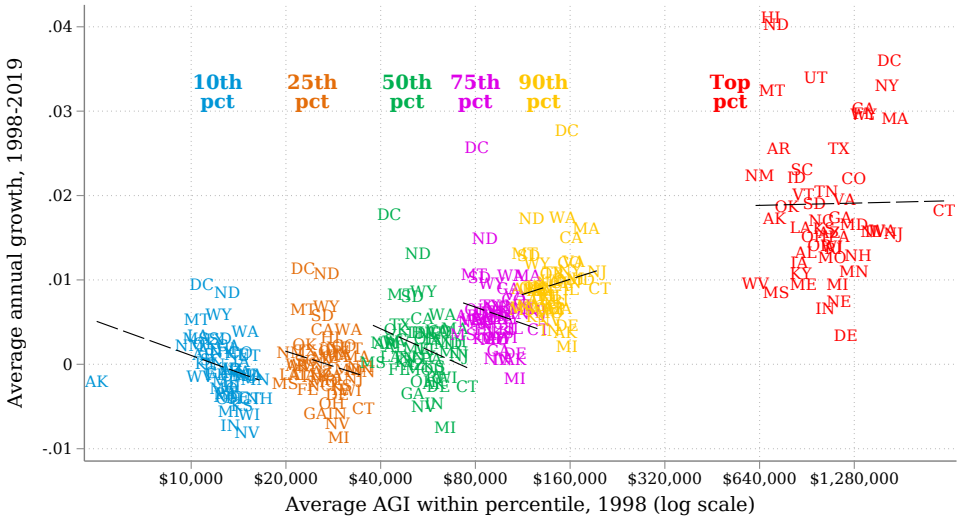
Figure 2: β-convergence at the mean in multiple income measures



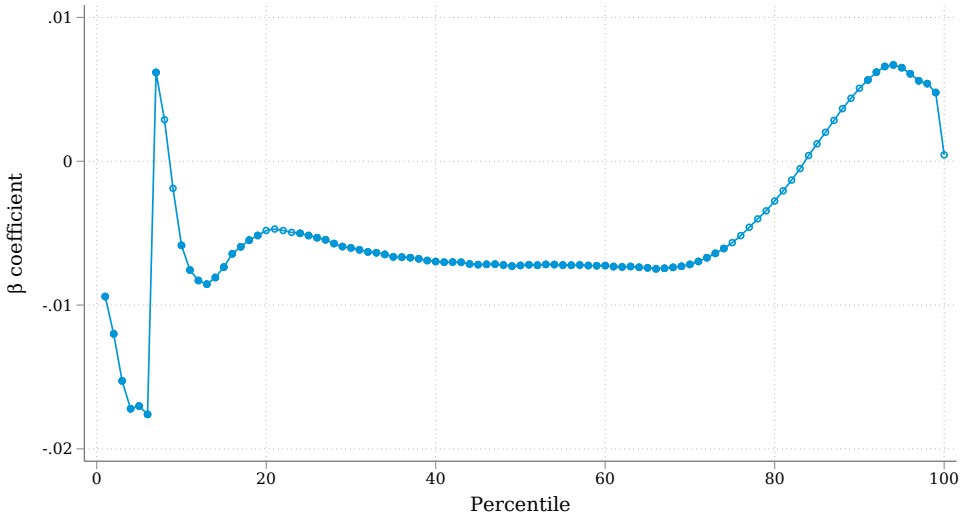
Source: Bureau of Economic Analysis, IRS Form 1040, and authors' calculations.
Note: Values are in 2019 dollars, adjusted for inflation using the CPI-U. The x-axis is in log scale. PCPI = per capita personal income. The adjusted gross income (AGI) plot is based on average AGI in tax units with positive AGI. Release authorization CBDRB-FY23-051.

Figure 3: β-convergence across the income distribution, 1998-2019

(a) Scatter plots for select percentiles



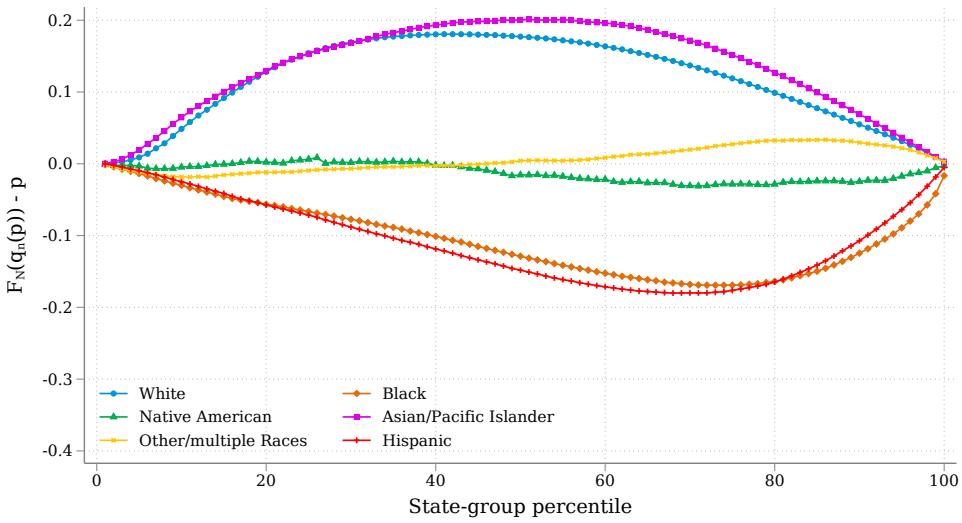
(b) β coefficients for all percentiles



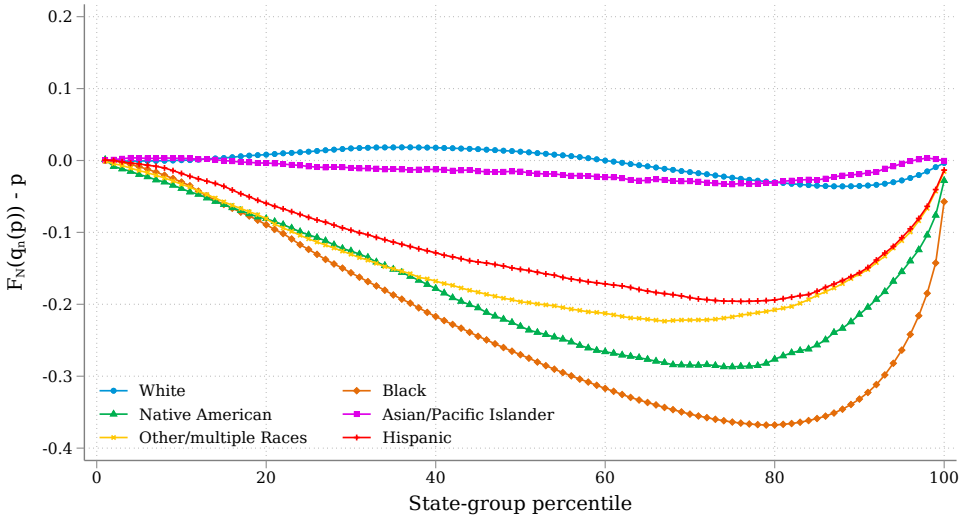
Source: IRS Form 1040 and authors' calculations.
Note: Adjusted gross income values are in 2019 dollars, adjusted for inflation using the CPI-U. The x-axis in panel (a) is in log scale. Analysis is at the tax unit level and limited to tax units with positive values of adjusted gross income. In calculating percentile cutoffs, tax units are weighted by size. Dashed lines in panel (a) depict the β coefficient from the regression of the form Average annual growth, 1998-2019 $= \alpha + \beta$log(1998 income level). Panel (b) depicts that coefficient for all percentiles. Filled points in panel (b) indicate statistical significance at at least the 10 percent level. Release authorization CBDRB-FY23-051.

Figure 4: Example Relative Regional Income Divergence Curves, 2019

(a) Connecticut



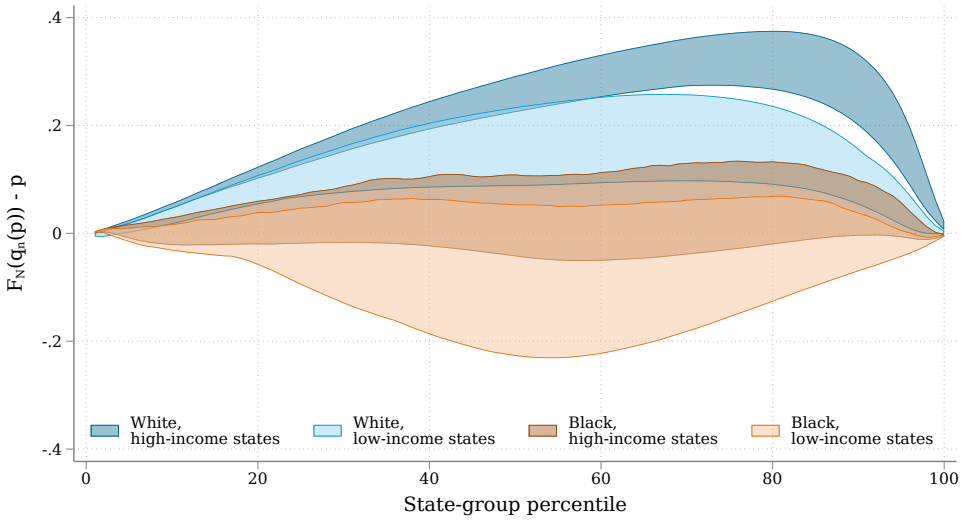
(b) Mississippi



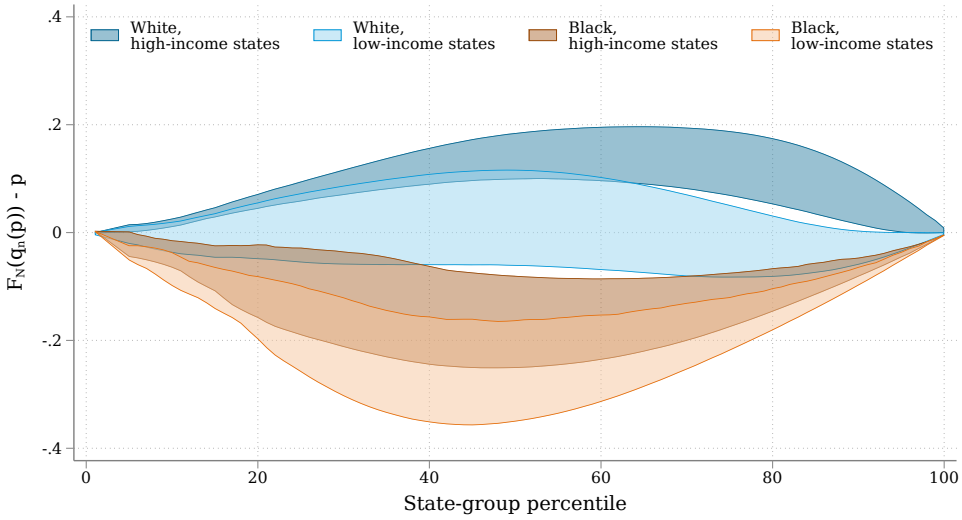
Source: IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: Relative regional income divergence curves constructed as in Section 4.2.1 for each race/ethnicity group using adjusted gross income (AGI) from tax units with positive values of AGI. Each person is assigned the AGI value of the tax unit to which they belong. Release authorization CBDRB-FY23-051.

Figure 5: Relative income by race in high- vs. low-income states

(a) 1998

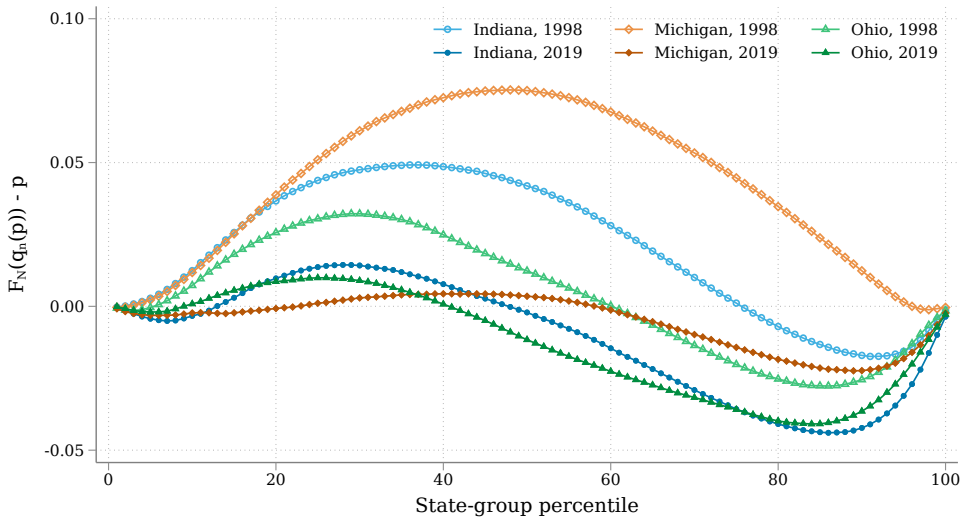


(b) 2019



Source: IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: Relative regional income divergence curves constructed as in Section 4.2.1 for each race/ethnicity group using adjusted gross income (AGI) from tax units with positive values of AGI. Each person is assigned the AGI value of the tax unit to which they belong. High-income states are the ten states (excluding the District of Columbia) with the highest personal income per capita in 1998. Low-income states are the ten states with the lowest personal income per capita in 1998. These groups are consistent across both panels of the figure. The plotted ranges depict the minimum and maximum value of the relative regional income divergence curve for each group at each percentile. Release authorization CBDRB-FY23-051.

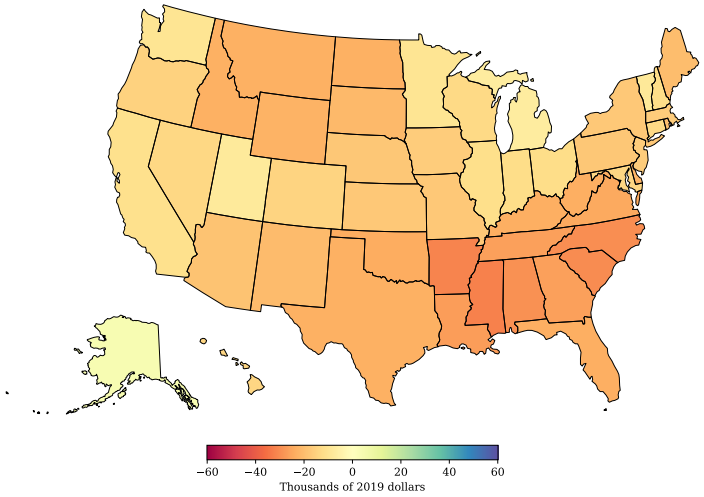
Figure 6: Relative income divergence among men, select states



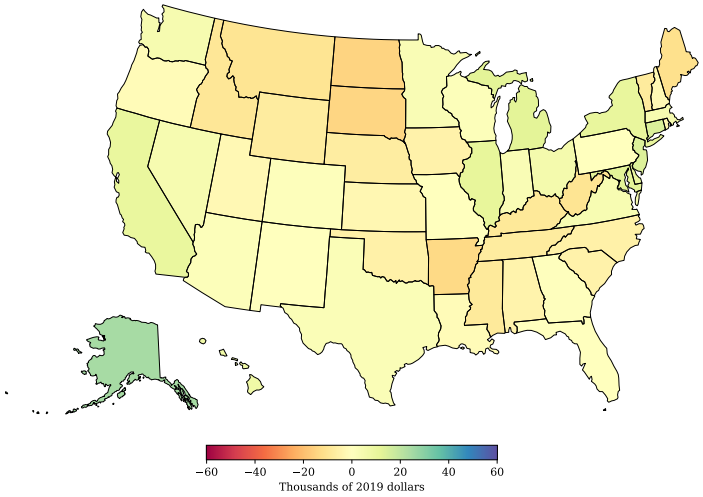
Source: Census Numident, IRS Form 1040, and authors' calculations.
Note: Relative regional income divergence curves constructed as in Section 4.2.1 for each gender using adjusted gross income (AGI) from tax units with positive values of AGI. Each person is assigned the AGI value of the tax unit to which they belong. Release authorization CBDRB-FY23-051.

Figure 7: Absolute divergence index by race and year, $\nu = 1$

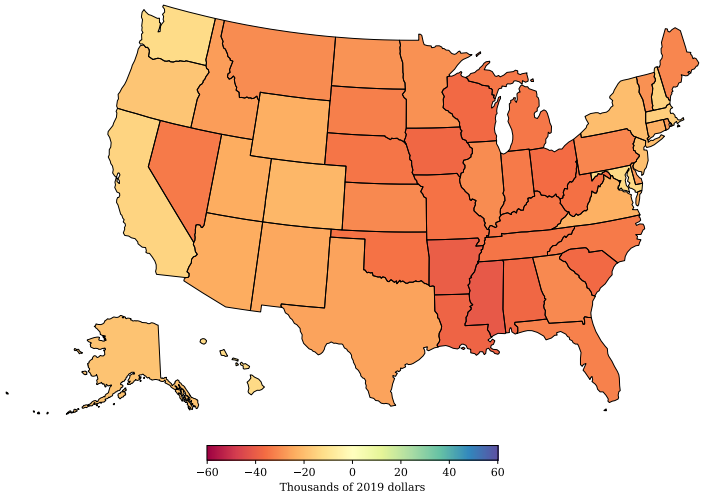
(a) Black, 1969



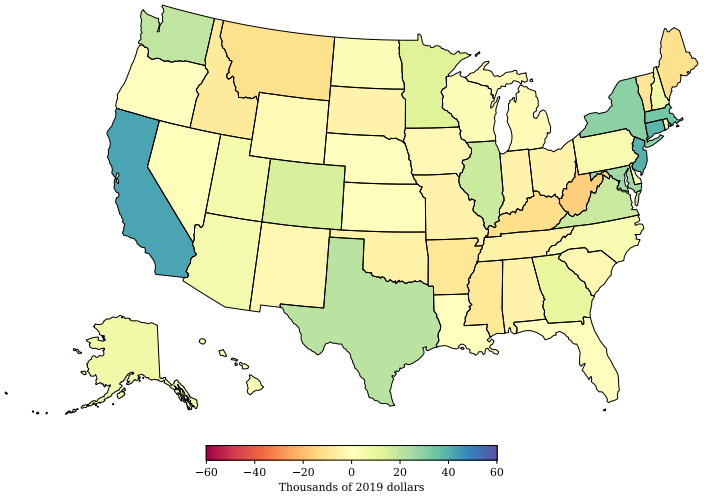
(b) White, 1969



(c) Black, 2019



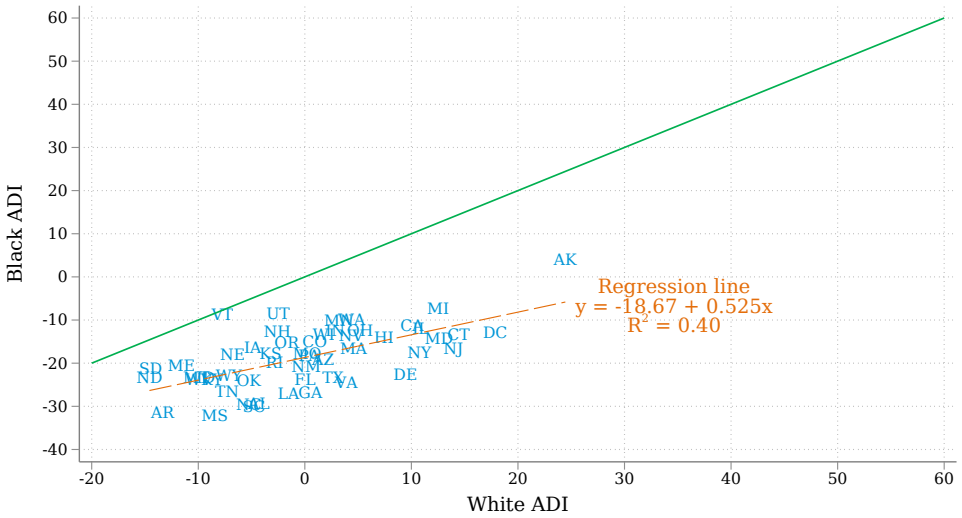
(d) White, 2019



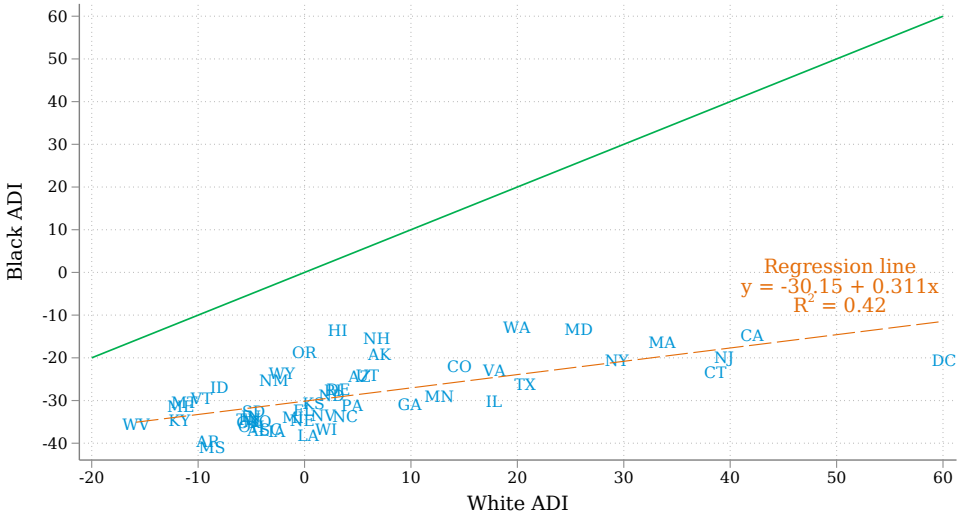
Source: Census Numident, IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: The absolute divergence index is created as in 4.2.2 from underlying absolute regional income divergence curves (ARIDCs) using Gini weights with parameter $\nu = 1$. The underlying ARIDCs are constructed as in Section 4.2.1 for each race/ethnicity group using wage and salary income from tax units with positive values of wage and salary income. Analysis is limited to prime-age people (ages 25-54). Each person is assigned the wage and salary income of the tax unit to which they belong. Release authorization CBDRB-FY23-051.

Figure 8: Absolute divergence index, Black vs. White, $\nu = 1$

(a) 1969

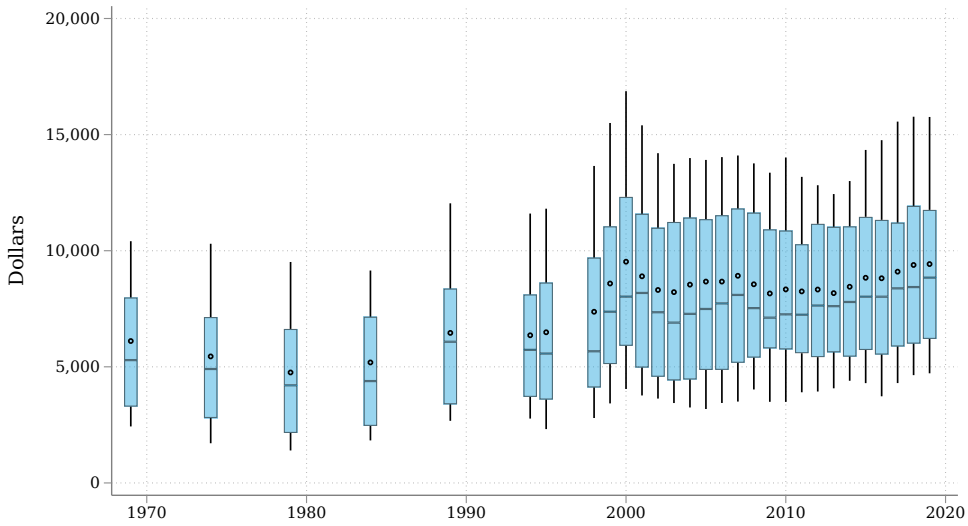


(b) 2019



Source: Census Numident, IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: Solid green lines show equality between the measures on the two axes. Dashed orange lines show regression lines associated with each scatter plot. The absolute divergence indexes are created as in 4.2.2 from underlying absolute regional income divergence curves (ARIDCs) using Gini weights with parameter $\nu = 1$. The underlying ARIDCs are constructed as in Section 4.2.1 for each race/ethnicity group using wage and salary income from tax units with positive values of wage and salary income. Analysis is limited to prime-age people (ages 25-54). Each person is assigned the wage and salary income of the tax unit to which they belong. Release authorization CBDRB-FY23-051.

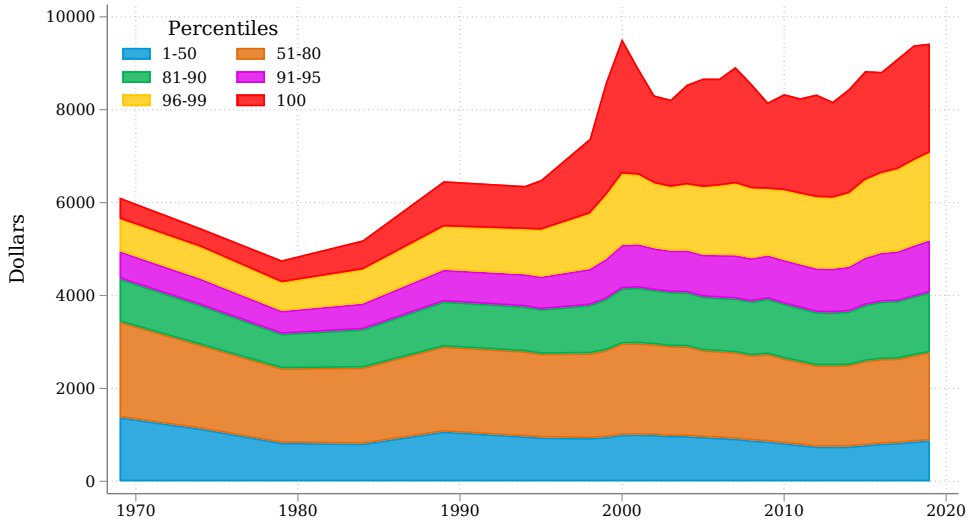
Figure 9: Absolute difference from national distribution, average across percentiles within states



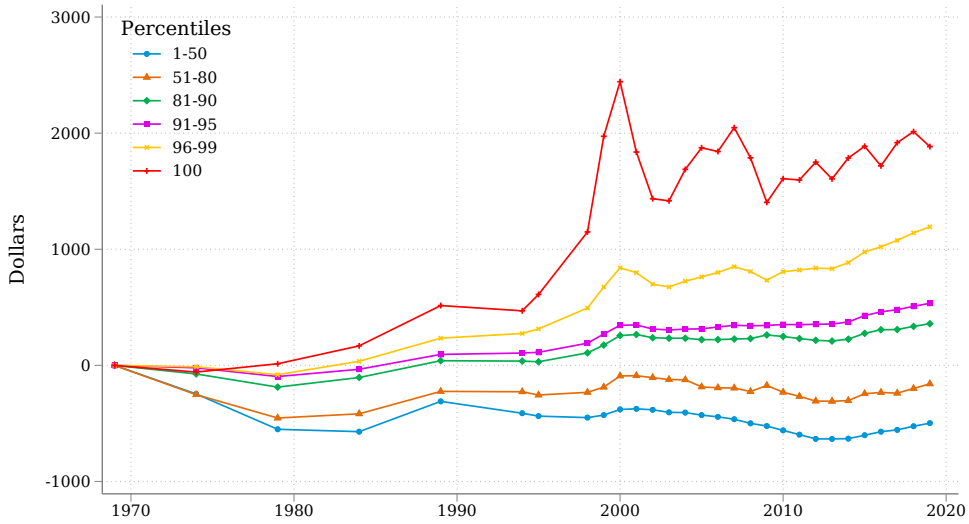
Source: IRS Form 1040 and authors' calculations.
Note: The figure shows key moments of the distribution of state average absolute differences from the national income distribution across all percentiles of the wage and salary income distribution, as in Equation 2. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of average absolute differences. The circles represent the mean; the blue box represents the interquartile range; the bar within the blue box represents the median; and the lines extending from the blue box reach the 10th (below) and 90th (above) percentiles. Release authorization CBDRB-FY23-051.

Figure 10: Decomposition of convergence across the distribution of income

(a) Contribution to AAD



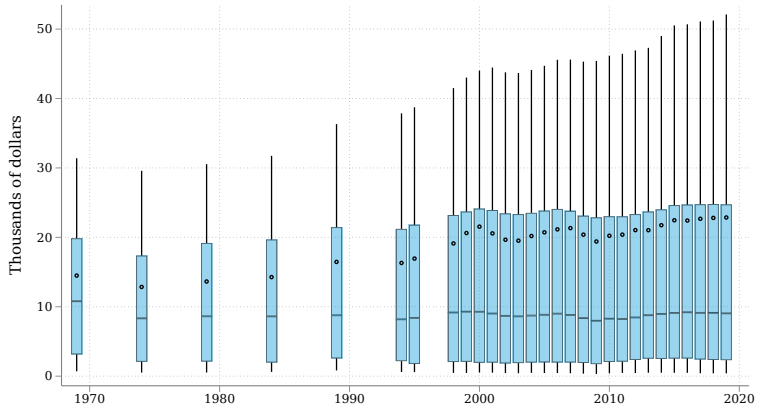
(b) Change in contribution from 1969



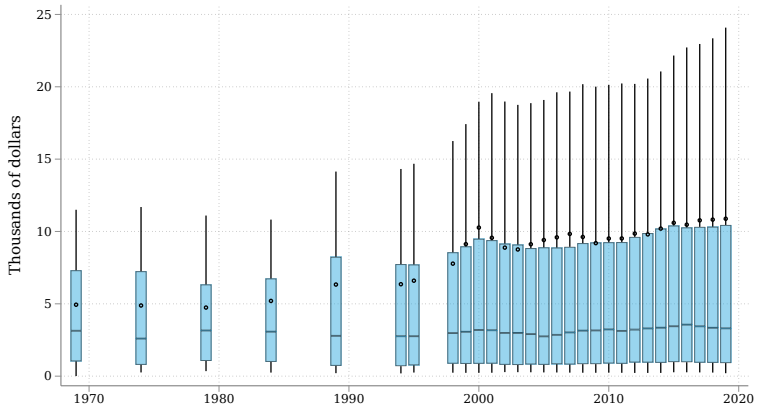
Source: IRS Form 1040 and authors' calculations.
Note: Figures plot the average absolute difference (AAD) across states from the national wage and salary income distribution, as in Equation 2, within the indicated percentile ranges. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of average absolute differences. The combined level in panel (a) gives the average across all percentiles. Panel (b) normalizes each group's contribution to zero in 1969. Release authorization CBDRB-FY23-051.

Figure 11: Absolute difference from national distribution, average across percentiles within states, by race

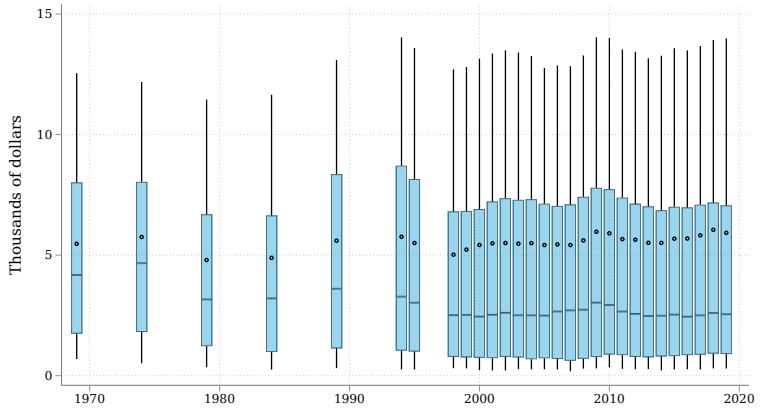
(a) Black vs. national overall distribution



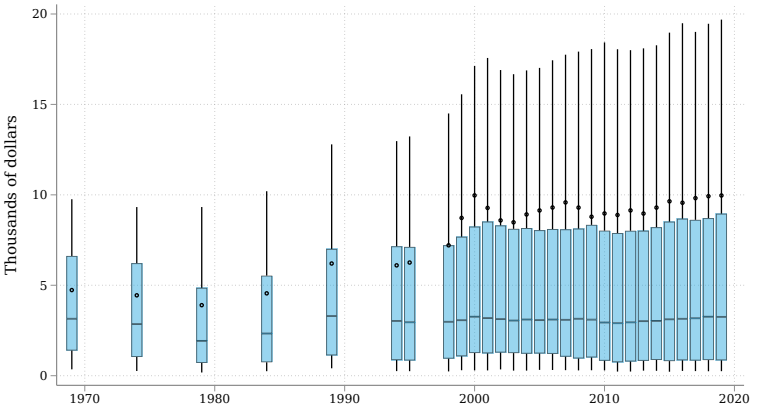
(b) White vs. national overall distribution



(c) Black vs. national Black distribution



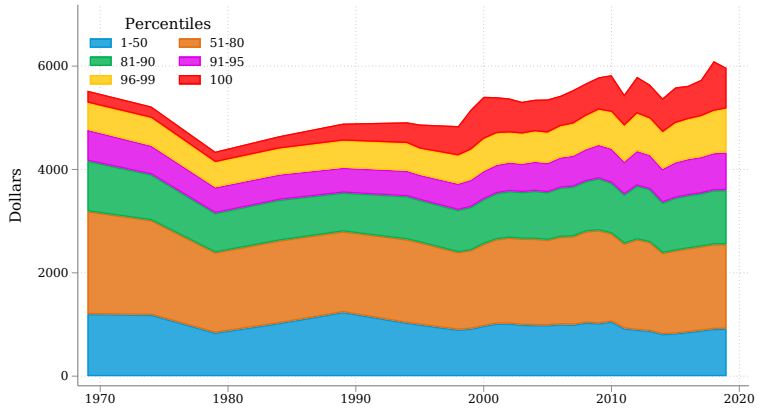
(d) White vs. national White distribution



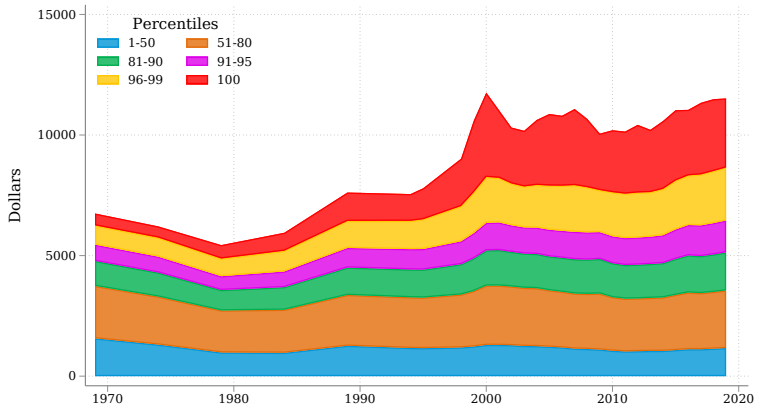
Source: IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: The figure shows key moments of the distribution of state average absolute differences from the national income distribution across all percentiles of the wage and salary income distribution, as in Equation 2. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of average absolute differences. In panels (a) and (b), the national distribution used to construct average absolute differences is the overall national income distribution, including all people regardless of race. In panels (c) and (d), the national distributions used are the national distributions within the corresponding race group (i.e. state-level White distributions are compared to the national White distribution). The circles represent the mean; the blue box represents the interquartile range; the bar within the blue box represents the median; and the lines extending from the blue box reach the 10th (below) and 90th (above) percentiles. Release authorization CBDRB-FY23-0136.

Figure 12: Absolute difference from national distribution, average across percentiles within states, by race

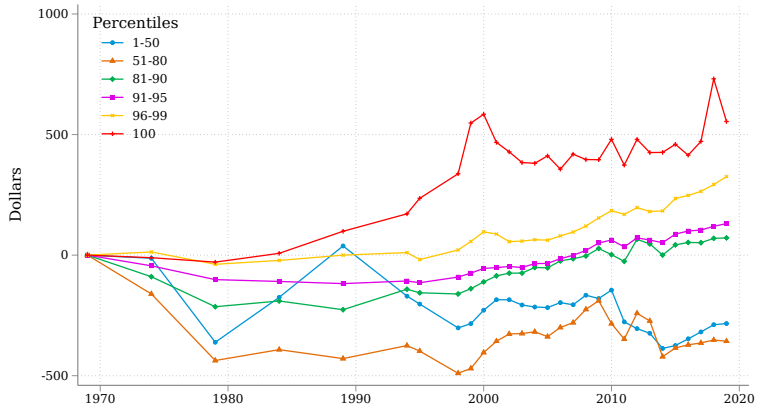
(a) Black, contribution to AAD



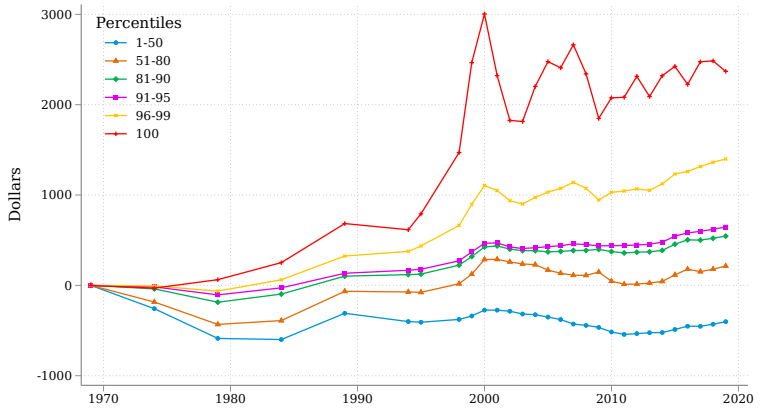
(b) White, contribution to AAD



(c) Black, change in contribution from 1969



(d) White, change in contribution from 1969

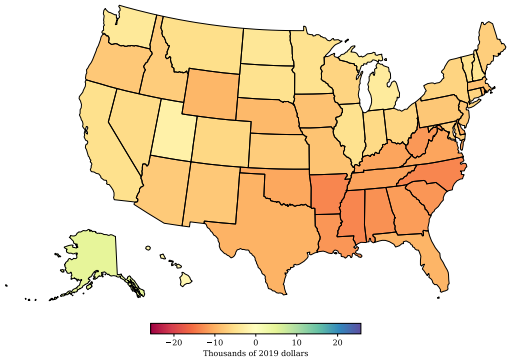


Source: IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: Figures plot the average absolute difference (AAD) across states from the national wage and salary income distribution, as in Equation 2, within the indicated percentile ranges. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of average absolute differences. The national distributions used to construct average absolute differences are the national distributions within the indicated race group (i.e. state-level White distributions are compared to the national White distribution). The combined levels shown in panels (a) and (b) give the average across all percentiles. Panels (c) and (d) normalize each group's contribution to zero in 1969. Release authorization CBDRB-FY23-0136.

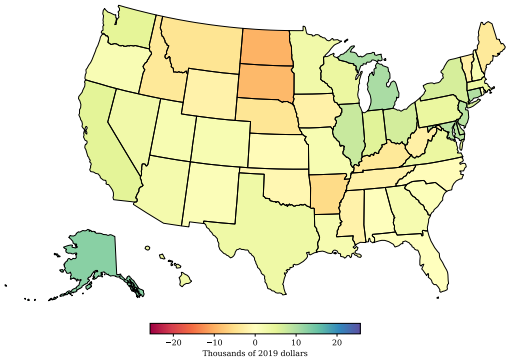
Appendix A Additional Figures

Figure A1: Absolute divergence index by race and year, $\nu = 5$

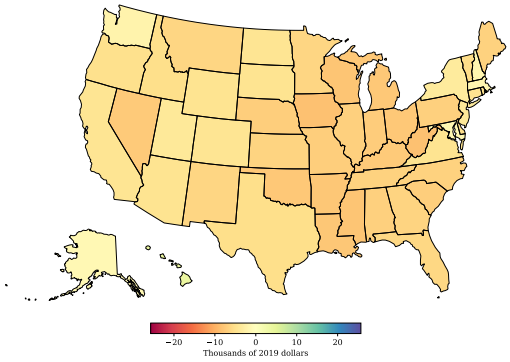
(a) Black, 1969



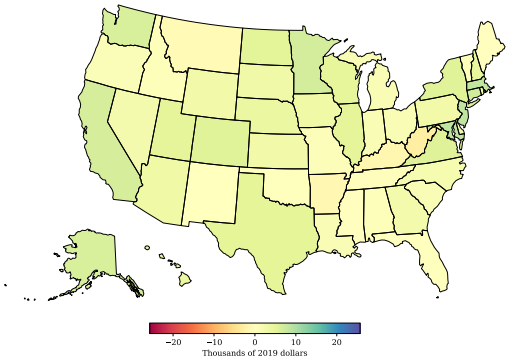
(b) White, 1969



(c) Black, 2019



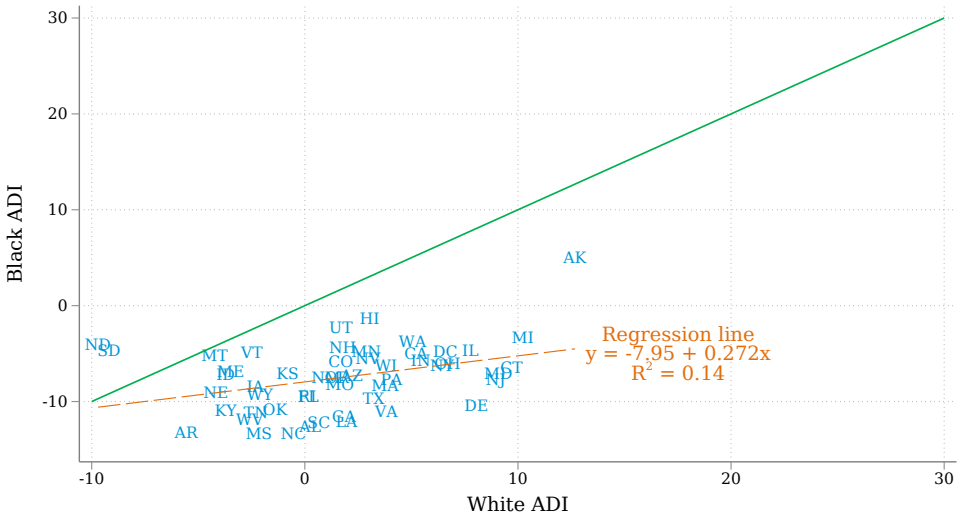
(d) White, 2019



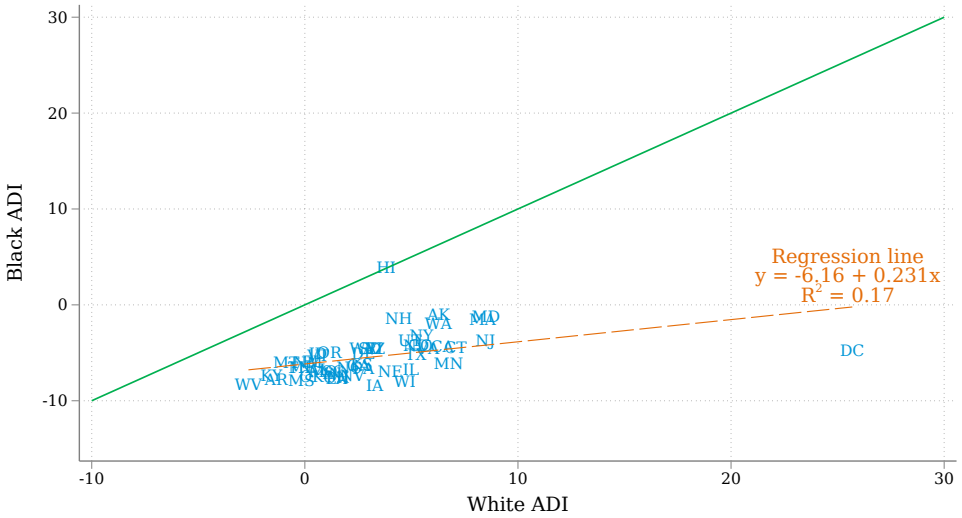
Source: Census Numident, IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: The absolute divergence index is created as in 4.2.2 from underlying absolute regional income divergence curves (ARIDCs) using Gini weights with parameter $\nu = 5$. The underlying ARIDCs are constructed as in Section 4.2.1 for each race/ethnicity group using wage and salary income from tax units with positive values of wage and salary income. Analysis is limited to prime-age people (ages 25-54). Each person is assigned the wage and salary income of the tax unit to which they belong. Release authorization CBDRB-FY23-051.

Figure A2: Absolute divergence index, Black vs. White, $\nu = 5$

(a) 1969

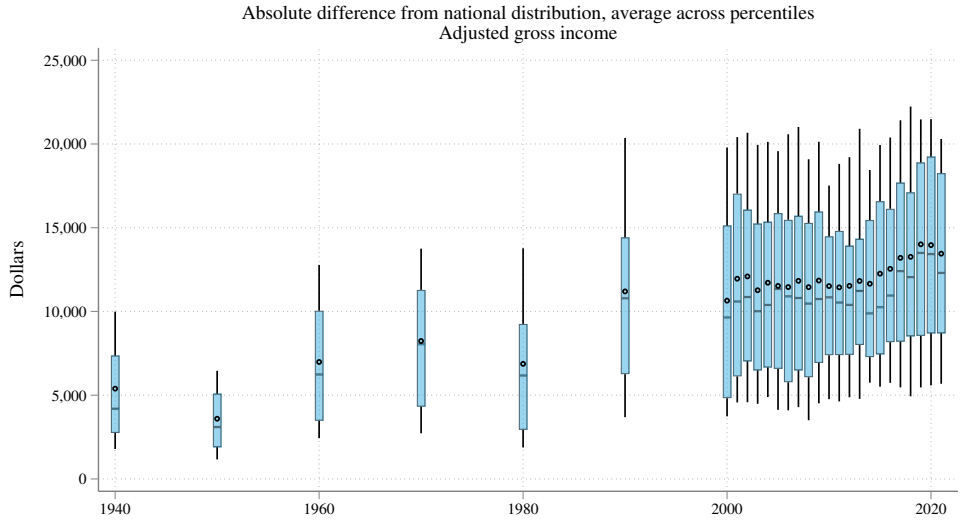


(b) 2019



Source: Census Numident, IRS Form 1040, Best Race and Ethnicity Administrative Records Composite file, and authors' calculations.
Note: Solid green lines show equality between the measures on the two axes. Dashed orange lines show regression lines associated with each scatter plot. The absolute divergence indexes are created as in 4.2.2 from underlying absolute regional income divergence curves (ARIDCs) using Gini weights with parameter $\nu = 5$. The underlying ARIDCs are constructed as in Section 4.2.1 for each race/ethnicity group using wage and salary income from tax units with positive values of wage and salary income. Analysis is limited to prime-age people (ages 25-54). Each person is assigned the wage and salary income of the tax unit to which they belong. Release authorization CBDRB-FY23-051.

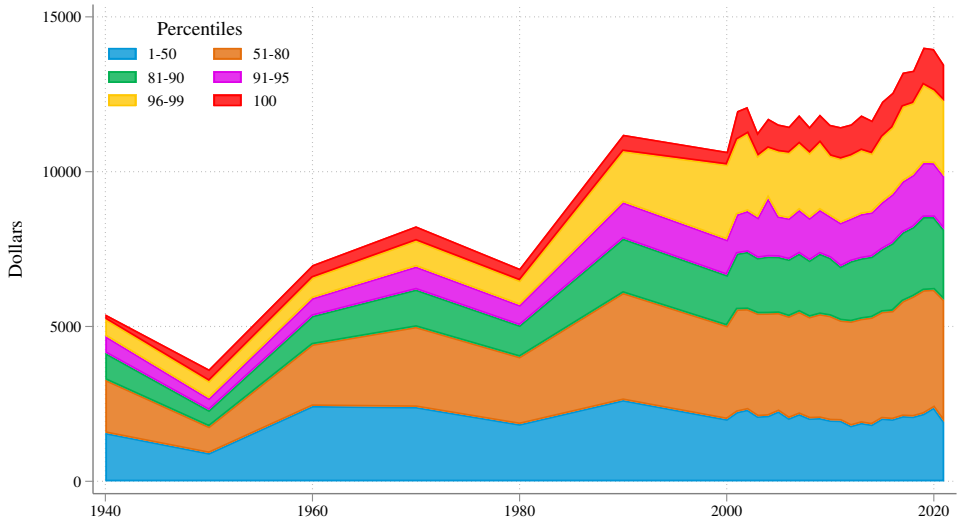
Figure A3: Absolute difference from national distribution, average across percentiles within states (IPUMS)



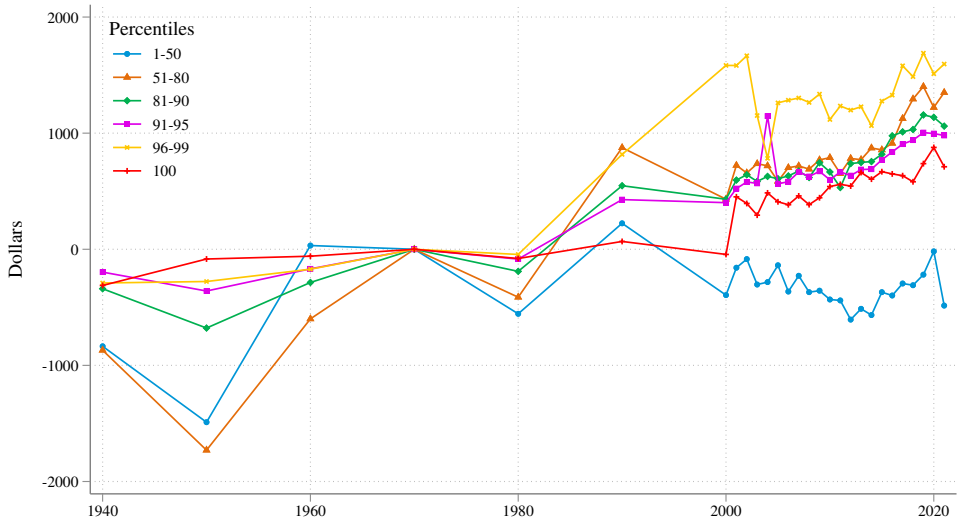
Source: IPUMS decennial census and American Community Survey data and authors' calculations.
Note: The figure shows key moments of the distribution of state average absolute differences from the national income distribution across all percentiles of the wage and salary income distribution, as in Equation 2. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of average absolute differences. The circles represent the mean; the blue box represents the interquartile range; the bar within the blue box represents the median; and the lines extending from the blue box reach the 10th (below) and 90th (above) percentiles.

Figure A4: Decomposition of convergence across the distribution of income (IPUMS)

(a) Contribution to AAD



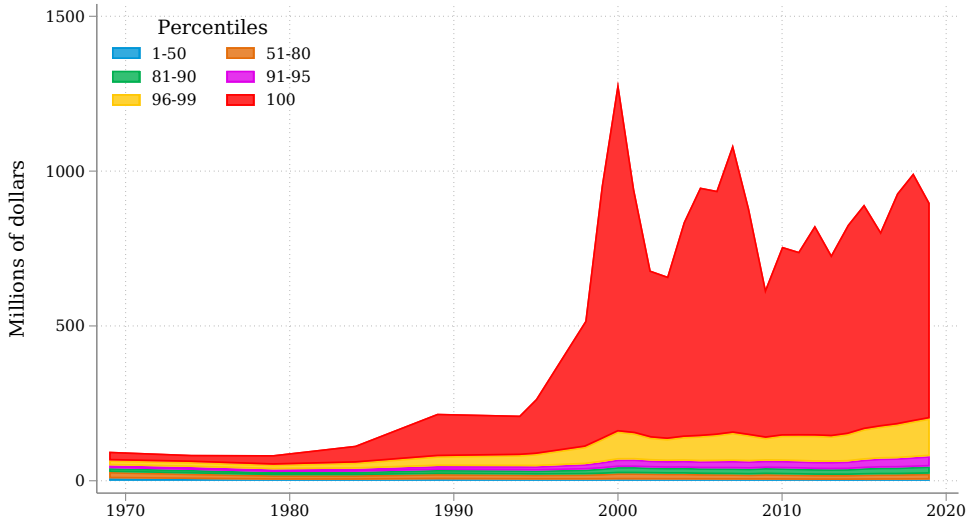
(b) Change in contribution from 1970



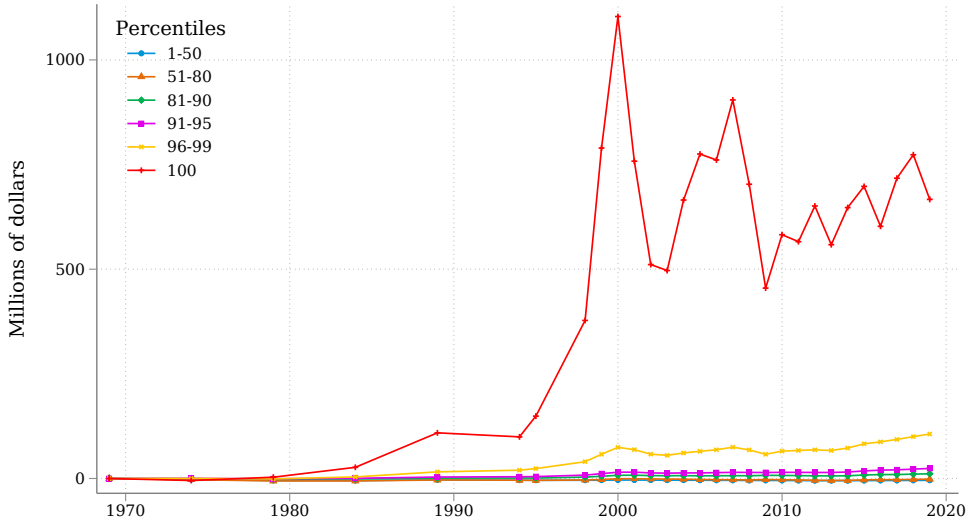
Source: IPUMS decennial census and American Community Survey data and authors' calculations.
Note: Figures plot the average absolute difference (AAD) across states from the national wage and salary income distribution, as in Equation 2, within the indicated percentile ranges. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of average absolute differences. The combined level in panel (a) gives the average across all percentiles. Panel (b) normalizes each group's contribution to zero in 1970.

Figure A5: Decomposition of convergence across the distribution of income, measured using mean squared difference

(a) Contribution to AAD



(b) Change in contribution from 1969



Source: IRS Form 1040 and authors' calculations.
Note: Figures plot the mean squared difference (MSD) across states from the national wage and salary income distribution, constructed analogously to Equation 2, within the indicated percentile ranges. State and national average incomes within percentiles are converted to 2019 dollars using the CPI-U prior to the construction of mean squared differences. The combined level in panel (a) gives the average across all percentiles. Panel (b) normalizes each group's contribution to zero in 1969. Release authorization CBDRB-FY23-0136.

Appendix B State Income Data Dictionaries

Three data files providing measures related to the analysis presented above for all states accompany this paper.

Measures are generated using the universe of 1040 tax returns with positive values of the income concepts in question. Percentiles are constructed to represent one percent of people (rather than tax returns) in each group. Each record is assigned its tax unit's value of each income measure.

Files are available here, as is a stand-alone version of this appendix. The files are

- state_curves.csv: This file provides ordinates of absolute and relative regional income divergence curves (see Section 4.2.1) for each year from 1998–2019 and for all states. In addition to curves for the full population of each state, it provides curves for subgroups defined by gender (male, female) and race/ethnicity (non-Hispanic White alone; non-Hispanic Black alone; non-Hispanic Asian alone, including Native Hawaiians and other Pacific Islanders; and Hispanic). Only curves for which every ordinate is based on at least 30 observations are reported.

- state_indexes.csv: This file contains aggregations of the absolute and relative regional income divergence curves using various Gini weights (see Section 4.2.2). These regional income divergence indexes were produced from the estimates in state_curves.csv after they were disclosed, so they are only available for groups for which absolute and relative regional income divergence curves are available.

- nat_dists.csv: This file provides information about the full-population national income distribution in each year. Estimates from state_curves.csv can be combined with estimates in this file to produce absolute measures of income in dollars or percentiles of the national income distribution.

The tables below describe the variables available in each file and explain what their values represent.

Table B1: Data dictionaries for state_curves.csv

Variable name and description	Value	Value definition
file: how are subgroups within states defined?	all	Estimates for the full population regardless of race/gender
	gender	Estimates for groups defined by gender
	racegrp	Estimates for groups defined by race/ethnicity
inctype: what income concept is used?	AGI	Estimates are based on tax unit adjusted gross income
	WSI	Estimates are based on tax unit wage and salary income
year: what year are estimates for?	1998–2019	Esimates are based on data from this tax year
state: what state are estimates for?	postal codes for 50 states and DC	Estimates are based on data from this state
groupid: what subgroup are estimates for?	0	For **file** = all, 0 = full population
	1	For **file** = gender, 1 = male
		For **file** = racegrp, 1 = non-Hispanic White
	2	For **file** = gender, 2 = female
		For **file** = racegrp, 2 = non-Hispanic Black
	4	For **file** = racegrp, 4 = non-Hispanic Asian (including native Hawaiian and other Pacific islanders)
	6	For **file** = racegrp, 6 = Hispanic
pct: what percentile are estimates for?	1–100	Estimates are based on data from this percentile
rridc: what is the average of $F_N(q_n(p)) - p$ for people in this year-state-group-percentile?	[min,max]	Average value of the relative regional income divergence curve (RRIDC) within state-group-percentile
aridc: what is the average of $q_N(p) - q_n(p)$ for people in this year-state-group-percentile?	[min,max]	Average value of the absolute regional income divergence curve (ARIDC) within state-group-percentile in nominal dollars

Table B2: Data dictionary for state_indexes.csv

Variable name and description	Value	Value definition
file: how are subgroups within states defined?	all	Estimates for the full population regardless of race/gender
	gender	Estimates for groups defined by gender
	racegrp	Estimates for groups defined by race/ethnicity
inctype: what income concept is used?	AGI	Estimates are based on tax unit adjusted gross income
	WSI	Estimates are based on tax unit wage and salary income
year: what year are estimates for?	1998–2019	Esimates are based on data from this tax year
state: what state are estimates for?	postal codes for 50 states and DC	Estimates are based on data from this state
groupid: what subgroup are estimates for?	0	For **file** = all, 0 = full population
	1	For **file** = gender, 1 = male
		For **file** = racegrp, 1 = non-Hispanic White
	2	For **file** = gender, 2 = female
		For **file** = racegrp, 2 = non-Hispanic Black
	4	For **file** = racegrp, 4 = non-Hispanic Asian (including native Hawaiian and other Pacific islanders)
	6	For **file** = racegrp, 6 = Hispanic
nu: what value of ν is used to create the Gini weights used to construct these estimates?	1–5	Gini weights are constructed using this value of ν
adi: what is the average value of the ARIDC within this year-state-group?	[min,max]	Average value of the ARIDC, in thousands of nominal dollars
rdi: what is the average value of the ARIDC within this year-state-group?	[min,max]	Average value of the RRIDC

Table B3: Data dictionary for nat_dists

Variable name and description	Value	Value definition
measure: what income concept is used?	AGI	Estimates are based on tax unit adjusted gross income
	WSI	Estimates are based on tax unit wage and salary income
year: what year are estimates for?	1998–2019	Estimates are based on data from this tax year
pct: what percentile are estimates for?	1–100	Estimates are based on data from this percentile
meaninc: what is the average income within this percentile?	[min,max]	Average income within this percentile in nominal dollars
cutoff: what is the maximum income value within this percentile?	[min,max]	Income value y such that **pct**% of people have income less than or equal to y, in nominal dollars