

PLEASE DO NOT CITE OR FORWARD. PRELIMINARY RESULTS ONLY.

## Hispanic and Asian Earnings Inequality and the Role of Labor Market Entrants and Immigrants

Randall Akee<sup>†</sup>, Maggie R. Jones<sup>‡</sup>, Sonya R. Porter<sup>‡</sup>, Emilia Simeonova<sup>\*</sup>

<sup>†</sup>Brookings Institution, UCLA, and NBER

<sup>‡</sup>U.S. Census Bureau

<sup>\*</sup>Johns Hopkins University and NBER

November 21, 2019

### Abstract

Economic inequality has been increasing in the U.S. over the past several decades. The contribution of purely economic factors, such as wage divergence within a generation over time, versus demographic and societal contributors, such as selective immigration and changes in the earnings potential of new generations entering the labor market, is not well understood. The distinction between different mechanisms driving inequality may be especially relevant for racial and ethnic groups that experience high rates of immigration and demographic change. Using confidential-use, individual-level Internal Revenue Service and U.S. Census data, we follow the earnings of Hispanics and Asians between the ages of 18–45 with panel data that spans the years 2005–2014. We examine the impact that labor market entrants and new immigrant arrivals within each group have on group earnings inequality. We show that labor market entrants and immigrants increase inequality for both groups.

**Keywords:** Earnings inequality, Hispanic, Asian, race, ethnicity, administrative data

**JEL classification:**

**Acknowledgments** Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Data in this report based on confidential data were reviewed and released by the Census Bureau Disclosure Review Board, approval CBDRB-FY2019-CES005-015, CDRB-FY19-370. Any errors are ours alone. Akee and Simeonova acknowledge funding from the Robert Wood Johnson Foundation. **Author contact information:** Akee: rakee@ucla.edu; Jones: margaret.r.jones@census.gov; Porter: sonya.rastogi.porter@census.gov; Simeonova: emilia.simeonova@gmail.com.

# 1 Introduction and Background

Economic inequality in the U.S. has been rising over the past 30 years (Piketty, Thomas and Saez, Emmanuel, 2003), and there is growing evidence that this trend persists overall and within race and ethnic groups (Akee et al., 2019; Bloome, Deirdre, 2014; Reardon, Sean F and Bischoff, Kendra, 2011; Snipp, C Matthew and Cheung, Sin Yi, 2016). Previous research attributed this increase in inequality to changes in the demand for skill from the 1960s onward (Juhn et al., 1993). Other explanations for increasing earnings inequality have focused on the role of the real minimum wage (Card and DiNardo, 2002) and skill-biased technological change, where new computing technology complements high-skilled workers and substitutes for low or medium-skilled workers (Autor et al., 2008). These studies examined demand-side determinants of earnings inequality. Due to data limitations or sample size restrictions, little work has been done to investigate potential supply-side contributors to these trends. For example, changes in the composition of labor market participants arising from demographic changes within race and ethnic groups, compounded by recent immigration trends, may affect within- and between-group earnings inequality.

Our analysis aims to explore supply-side explanations of earnings inequality by linking annual Internal Revenue Service (IRS) W-2 and 1099 tax records to various American Community Surveys (ACS); this allows us to follow an individual’s earnings experience for the time period 2005 to 2014. It also allows us to identify new immigrants to the U.S., which in turn permits an identification of two groups of labor market entrants: those who enter via recent immigration and those who enter from resident non-employment (either aging into the workforce, departing education at different levels, or simply reentering after a period of unemployment). We first aggregate the individual data to produce race and ethnic group-level inequality measures over time. Then we separately map out the earnings inequality of the continuously employed, the U.S. labor market entrants, and the recent immigrant arrivals over the decade. This analysis allows us to fully disentangle earnings inequality within minority groups to assess the separate contributions of domestically produced labor market entrants and of recent immigrants, which in turn sheds light on the extent to which these two types of new workers are substitutes or complements in the U.S. labor market.

In cross-sectional analyses using aggregate data by ethnic and race groups (non-Hispanic white, Hispanic, Asian), we find that each group exhibits increasing earnings inequality. However, when we disaggregate the race and ethnic groups into three labor-market subgroups—a continuously employed panel, U.S. labor market entrants, and new immigrant workers—we find that the growth in earnings

inequality for all three groups is flat for the continuously employed. In general, labor market entrants and new immigrants exhibit higher levels of earnings inequality than established workers, which explains the upward trend in inequality observed in repeated cross-section data. Meanwhile, patterns of inequality differ between minority groups, with new Asian and non-Hispanic white workers displaying stable or upward trends in inequality and Hispanics revealing flat trends.

Our analysis focuses on three race and ethnic groups—Hispanics, Asians, and non-Hispanic whites (who form our comparison group). The first two groups comprise the largest flow of immigrants to the U.S. in recent years and so may contribute disproportionately to differences in earnings inequality in analyses based on cross-sectional data that do not distinguish between established workers and labor market entrants. Additionally, non-immigrant labor market entrants from these two groups display considerable human-capital heterogeneity and thus may substantially change the overall inequality profile of their group upon entering the U.S. workforce. Finally, Asians and Hispanics have the highest growth rates as percent of the total population in the U.S. and thus will have a large impact on the composition of future labor market entrants; new projections indicate that the population of Asians in the U.S. will increase by 128 percent between 2014 and 2060, while the population of Hispanics in the U.S. will increase by 114 percent over the same period (Colby and Ortman, 2017). These projected increases are much larger than for any other of the major race and ethnic categories.

## 2 Data

### 2.1 Sample Selection and Linkage

In this analysis we link confidential-use individual records from the ACS to IRS W-2 and 1099 forms. We create a novel panel data set that follows the earnings of the Hispanic ethnic group and the Asian and non-Hispanic white race groups in the U.S. over 2005–2014. Our analysis assesses the growth in inequality between groups through a year-by-year comparison of within-group Gini coefficients, estimated overall and separated into continuously-employed and labor market entrants.

Our process for linking administrative records starts with the 2005 ACS. This survey is conducted annually on a representative sample of approximately 2-3 percent of the U.S. population. Individual-level records are assigned a protected identification key (PIK), which is unique across individuals and based on a person’s name, birth date, address, and social security number. The PIK is then used to match individuals to their IRS Form W-2 or 1099 records for each year from 2005 to 2014. We use both survey weights and inverse probability weights to account for any potential biases in the PIK

assignment and linking processes for the analyses that follow. In previous research [Akee et al. \(2019\)](#) we note that PIK assignment and linking to IRS data results in a data set that is skewed towards documented individuals, those employed in the formal labor force, and those with higher educational attainment.

In Appendix Table [A1](#) we provide the number of observations used in the first set of descriptive analyses for Hispanics and non-Hispanic whites from the 2005 ACS. There are approximately 1,319,000 Hispanics and non-Hispanic whites in the 2005 ACS who are between the ages of 18-45 and included in the initial sample; we are able to assign PIKs to 91 percent of those observations.<sup>1</sup> In the next row, we show the number of PIKed observations from the ACS found in the IRS W-2 or 1099 data, which is 988,000 or about 83 percent of the observations. This is the size of the final sample that we use in the analysis of Hispanic earnings inequality.

In the next two columns, we show the number of observations used in the second set of descriptive analyses for Asians and non-Hispanic whites starting with all the individuals identified in the ACS 2005 ages 18-45. There are approximately 1,295,000 Asian and non-Hispanic white individuals; we are able to assign PIKs to approximately 92 percent of those people or about 1,193,000 observations. In the next row, we show that 985,000 people were identified in the IRS W2 and 1099 data or approximately 83 percent of the observations with PIKs. This is the size of the final sample that we use in the analysis of Asian earnings inequality.

Using the annual earnings data, we create a panel data set of earnings for Asians, Hispanics and non-Hispanic whites that follows the same individuals (aged 18-45) over time. It is important to note that the panel is fixed at 2005; it follows the same individuals over time and examines the trajectory in earnings inequality for those who remain in the panel. The panel data includes both native-born and earlier immigrant arrival cohorts. An individual may exit the panel, but she needs to have at least two consecutive years of earnings data from 2005 onward to be included in the panel.

## 2.2 Labor Market Entrants and Recent Immigrants

The main panel data set begins in 2005 and follows the same group of individuals over time; it does not include any new entrants to the labor market, but it does allow for individuals to leave the labor market due to early retirement, death, or any other reason. One of the main advantages of using the confidential-use administrative data is that it allows us to identify recent immigrant workers and

---

<sup>1</sup>Note that these sample sizes have been rounded according to U.S. Census Disclosure Review Board rules to ensure confidentiality of the data.

U.S. labor market entrants. We add in U.S. labor market entrants from these three groups—Asians, Hispanics and non-Hispanic whites—in our analysis to examine the role that they play in observed earnings inequality separate from new immigrant workers.

Specifically, in each year from 2006 onward we include U.S. labor market entrants (which includes first-time labor market entrants as well as those who are returning to the labor force) as identified by their earnings (from a W-2 or 1099) in the current year and the absence of earnings in the previous year. We follow the labor market entrants from the year in which they enter the labor market and all subsequent years in the data set; furthermore, we retain all additional cohorts of labor market entrants for each subsequent year. We allow individuals to enter this sample in any year from 2006 onward and we restrict their ages to be the same as in the panel data. This allows us to capture U.S. labor market entrants who are entering the labor market for the first time, such as graduates from high school, college, or graduate school. It also allows us to include individuals who were out of the labor force previously and are returning to the labor force within the range of years examined.

As a separate group, we add in new immigrant workers to the U.S., including individuals who arrived in 2006 or later. We capture new immigrants via additional waves of the ACS (2006–2014), where recent immigrants are identified from a question about their year of arrival. We define a new immigrant as one who arrived in the year of the survey or the year before (thus the restriction to start in 2006). Earnings for new immigrants are also captured via W-2s and 1099s, making the results of our analysis applicable to immigrants who enter the formal U.S. labor market. Our ACS data support the conclusion that Hispanics and Asians represented the two largest immigrant flows to the U.S. in recent years.

Table 1 shows the relative proportions by our three race and ethnic groups from the 2006 ACS public-use data (Ruggles et al., 2019). In this table we show the proportion of each group for the continuously employed (panel), the labor market entrants and the recent immigrants. The table illustrates the relative differences in composition and relative proportions of each type of observation by race or ethnic group. Labor market entrants and new immigrants are proportionately larger amounts of the overall Asian and Hispanic groups than for non-Hispanic whites. This compositional difference is likely a major factor in the cross-sectional increase in earnings inequality over time for Asians and Hispanics relative to non-Hispanic whites.

**Table 1: Proportions by Race and Ethnic groups and Employment Type**

	NHW	Asian	Hispanic
Panel	87.12	79.06	79.07
Entrants	12.84	20.08	20.69
Immigrants	0.03	0.86	0.24

Source: 2006 American Community Survey, [Ruggles et al. \(2019\)](#)

### 3 Gini Coefficients

We calculate Gini coefficients for each year of our data for each ethnic and race group; then we calculate separate Gini's for each year by each subgroup defined by race or ethnicity and worker type. Thus each subgroup coefficient can be considered the contribution to inequality that each subgroup brings to the overall rate in each year. A higher coefficient value indicates a more unequal profile of workers within the subgroup.

In Figure 1 we provide the Gini coefficients by year for the full sample of labor market participants in the ACS who are non-Hispanic whites, Asians, and Hispanics. The plots are based on the full data for all three groups, including established workers, U.S. labor market entrants, and new immigrant workers. We compute the Gini coefficients for each demographic group in every year based on this cross-sectional sample. The plots derived from this cross-sectional analysis reveal interesting trends over time. Asians have consistently higher Gini coefficients than the other two groups, and the gap widens over the Great Recession, showing both increasing inequality within the Asian sample and increasing inequality relative to non-Hispanic whites and all Hispanics. Hispanics start out with the lowest inequality in 2005, but catch up with non-Hispanic whites over the great recession. By the end of the observation window, inequality among Hispanics is at the same level as non-Hispanic whites. Thus there is evidence that inequality in the cross-section of Asian and Hispanic earners rose over this 10-year period while it stayed roughly the same for non-Hispanic whites.

While higher Gini coefficients may indicate greater heterogeneity in skill level and returns to skill within a group, these level differences in inequality in 2005 align with the average earnings by group in 2006. Asians have the highest average income at \$32,000 and the highest Gini coefficients. Non-Hispanic whites have average earnings of \$29,500. Finally, Hispanics have the lowest average earnings in 2006 at \$17,700 and the lowest level of earnings inequality.<sup>2</sup>

In Figure 2 we present separate plots for the three subgroups of earners by Hispanic or Asian origin, where non-Hispanic whites form comparison groups. The panel set of observations starts in 2005,

<sup>2</sup>Average earnings calculated from 2006 American Community Survey public use data.

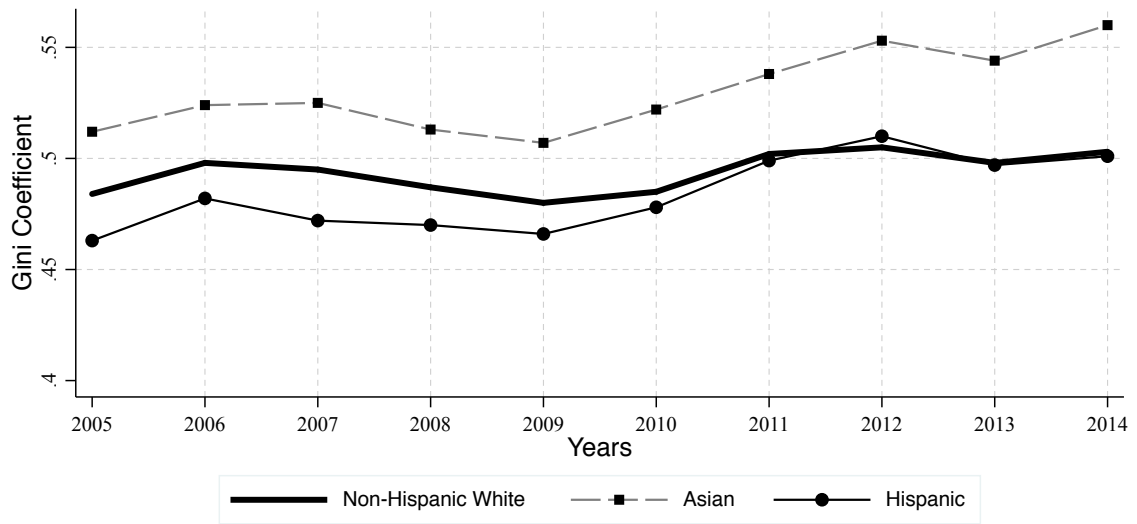
and those individuals are followed until 2014. The sample is selected on having at least 2 consecutive years of earnings records starting in 2005. U.S. labor market entrants are then included in a separate subgroup for the minority group in question and the non-Hispanic white comparison. The addition of the final subgroup—new immigrant workers—is handled in a similar manner.

The top panel in Figure 2 presents this dis-aggregation analysis for Asians, while the bottom panel shows the analysis for Hispanics. An assessment of the overall patterns reveals several novel findings. First, inequality is higher for both U.S. labor market entrants and new immigrant workers than the established workers in all groups. For each of the three demographic groups we find that the Gini coefficient ranges between 0.55 and 0.8 for new workers, while it is consistently below 0.5 for the panel observations. Non-Hispanic white immigrants experience a sharp increase in inequality over the Great Recession relative to U.S. labor market entrants (and the panel of established earners) that persists until at least 2014. However, they are a small proportion of the non-Hispanic white group overall (see Table 1) at only 0.03 of that group total. Second, new immigrants have lower Gini coefficients than U.S. labor market entrants for Asians and Hispanics, which is the opposite of what we find for non-Hispanic whites – suggesting that Hispanic and Asian immigrants are more homogeneous than the comparable set of non-Hispanic white immigrants over this time period.

In comparing patterns by the two minority groups, we see that the Asian panel of earners have higher levels of inequality than the non-Hispanic white panel, as does the group of Asian U.S. labor market entrants in comparison with non-Hispanic white entrants. However, the two entrant groups show similar patterns over time, although Asian entrants appear to experience an upward drift in inequality beginning in 2010. Meanwhile, although non-Hispanic white immigrants have a higher Gini than Asian immigrants in all years, both measures display an upward trend over time (for example, Asian immigrant inequality increases from about 0.625 to nearly 0.7). In tandem with evidence on the increasing share of high earnings for Asians (Akee et al., 2019), this result suggests a trend toward more higher-earning Asians in immigrant inflows.

The Hispanic panel reveals a different pattern. Each Hispanic subgroup has lower levels of inequality than its non-Hispanic white comparison in each year (with the possible exception of 2006 and 2008 for entrants). In addition, inequality trends for every Hispanic group are essentially flat, although there is wide variation in coefficient estimates for new immigrant workers that appears to correspond with the Great Recession and subsequent recovery. Overall, labor market entrants and new immigrants appear to reinforce the level differences we observe in the panel of established earners across non-Hispanic whites and Hispanics.

Figure 1: Gini Coefficients for Cross Section Data for Asian and Hispanics

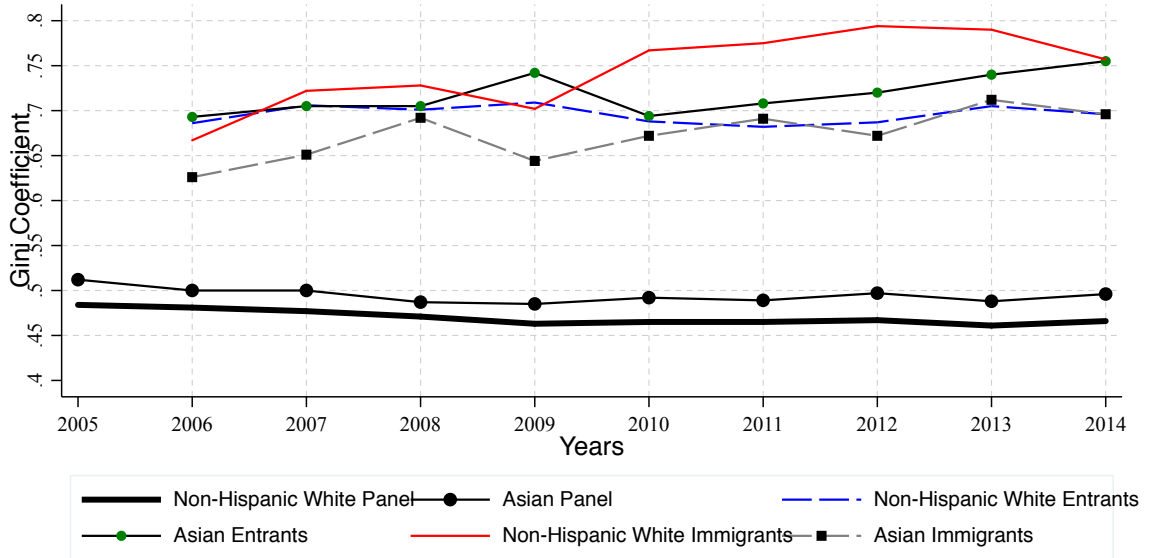


Source: 2005–2014 ACS and IRS 2005–2014 W-2 and 1099 data. Note: Numbers have been rounded to comply with the Census Bureau’s disclosure-avoidance guidelines. Data approved for dissemination by CBDRB-FY2019-CES005-015.

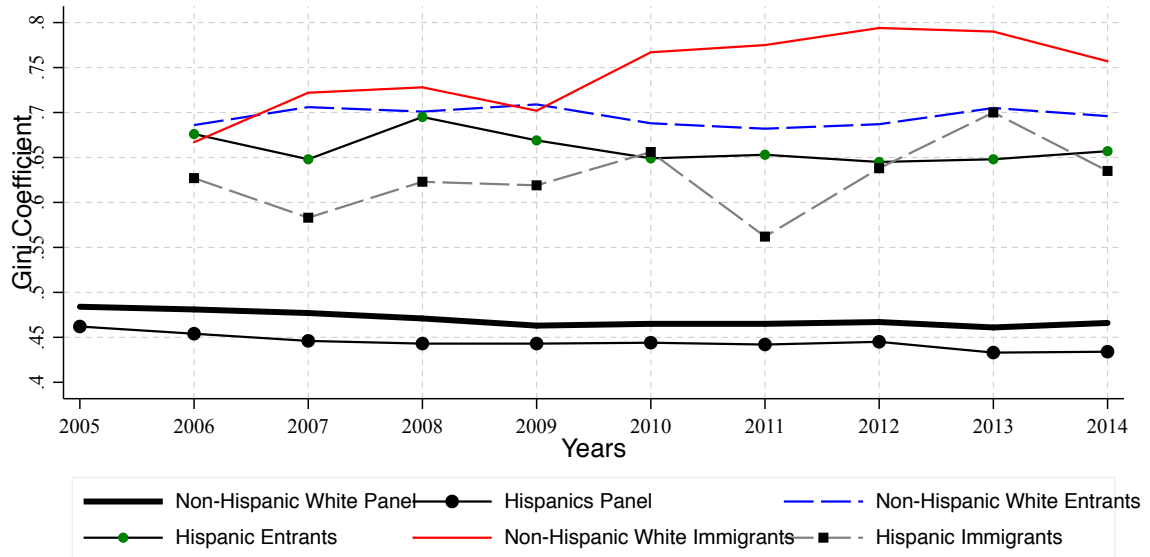


Figure 2: Gini Coefficients for Panel, Labor Market Entrants and New Immigrants

Panel A: Asian Panel, Labor Market Entrants and New Immigrants



Panel B: Hispanic Panel, Labor Market Entrants and New Immigrants



Source: 2005–2014 ACS and IRS 2005–2014 W-2 and 1099 data. Note: Numbers have been rounded to comply with the Census Bureau’s disclosure-avoidance guidelines. Data approved for dissemination by CBDRB-FY2019-CES005-015.

## 4 Measures of Human Capital

In Figure 2 we show that there are differences in inequality by race or ethnic group depending upon whether one examines the continuously employed, the U.S. labor market entrants, or the new immigrant workers. In Table 2 we show means of educational attainment for each of these groups from the 2006 ACS.<sup>3</sup> We report educational attainment for the three race and ethnic groups by their three subgroups in the nine columns in the table. We report average educational attainment by group in 2006, which aligns with the beginning of our observation period and prior to the Great Recession.

We find that, indeed, there are notable differences in educational attainment across these different race and ethnic groups and each of the labor market subgroups. Hispanic U.S. labor market entrants and recent immigrant workers have the lowest levels of human capital; a large proportion (83 percent to 85 percent) has only a high school degree or less. This contrasts with the sample of established Hispanic workers, who have higher levels of educational attainment than the entrants and immigrant groups.

We find slightly different results for Asians and non-Hispanic whites. Non-Hispanic white immigrants are better educated than the continuously employed workers and U.S. labor market entrants, suggesting increasing positive selection on education among non-Hispanic white immigrants (this is consistent with their observed drift upward in inequality). Non-Hispanic white labor market entrants who are not recent immigrants have the weakest education profile in this demographic group. Asian immigrants have a similar education profile to established Asian workers; on average both of those groups have higher educational attainment than Asian U.S. labor market entrants. This again suggests that there is strong positive selection on education among Asian immigrants, which is not the case for the Hispanic group.

Overall, the differences in educational attainment between established earners, U.S. labor market entrants, and recent immigrants across the three demographic groups appear to at least partially contribute to the different patterns of inequality documented in Figures 1 and 2. Changes in the composition of the labor force, especially in racial and ethnic groups that experience dynamic changes due to labor market entrants and immigration flows, are an important component in the investigation of inequality trends over time.

---

<sup>3</sup>The 2006 ACS data used in Table 2 includes individuals employed in the informal labor market, so the estimates of human capital attainment may be downward biased than for our more restricted, formal-labor-market-only data set.

**Table 2: Human Capital Proportions for Race and Ethnic groups**

	Panel			Entrants			Immigrants		
	NHW	Asian	Hispanic	NHW	Asian	Hispanic	NHW	Asian	Hispanic
HS or Less	0.43	0.27	0.74	0.59	0.37	0.83	0.45	0.32	0.85
Some College	0.27	0.21	0.16	0.22	0.24	0.11	0.18	0.16	0.04
College or More	0.29	0.52	0.08	0.17	0.37	0.04	0.36	0.51	0.06

Note: Figures above show the proportions of each column in each of the educational categories (shown in three rows); amounts may not sum to one due to rounding. Source: 2006 American Community Survey, [Ruggles et al. \(2019\)](#)

## 5 Conclusion

We construct a novel data set based on individual data on earnings, employment and immigration status to study the evolution of earnings inequality among Asians, Hispanics, and non-Hispanic whites over the past decade. Asians and Hispanics are two minority groups that have experienced rapid growth due to immigration and demographic change. To our knowledge, this is the first study to disentangle the contributions to within-group inequality of different types of new workers—entrants from the pool of potential workers from the group in the U.S. and recent immigrants from the same group. We also examine how differences in the relative proportion of these labor market subgroups in different race and ethnic groups contribute to observed cross-sectional earnings inequality. This supply-side examination provides additional potential explanations for observed increases in earnings inequality for specific groups over time.

Within the group of established workers, minority groups and non-Hispanic whites appear on similar inequality paths: we do not observe any substantial convergence or divergence between the three race and ethnic groups. Thus, even during a dynamic period such as the Great Recession, inequality trends across groups of established workers are stable. However, in cross-sectional analyses that cannot distinguish between different types of labor market participants, it is particularly important to recognize the contribution of U.S. labor market entrants and new immigrant workers. Their impact on average inequality and earnings mobility estimates is likely to be non-trivial, especially in racial and ethnic groups that experience periods of demographic growth and immigration. We find substantial variation in inequality depending on the race or ethnic group examined and the labor market subgroup in question. We posit that these differences are likely related to human capital differences—in particular differences in average educational attainment—which in turn has implications regarding the increasing heterogeneity of workers from each group and their differential returns in the labor market.

## References

- Akee, R., M. R. Jones, and S. R. Porter (2019). Race matters: Income shares, income inequality, and income mobility for all us races. *Demography* 56(3), 999–1021.
- Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics* 90(2), 300–323.
- Bloome, Deirdre (2014). Racial Inequality Trends and the Intergenerational Persistence of Income and Family Structure. *American Sociological Review* 79(6), 1196–1225.
- Card, D. and J. E. DiNardo (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of labor economics* 20(4), 733–783.
- Colby, S. L. and J. M. Ortman (2017). Projections of the size and composition of the us population: 2014 to 2060: Population estimates and projections.
- Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of political Economy* 101(3), 410–442.
- Piketty, Thomas and Saez, Emmanuel (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics* 118(1), 1–41.
- Reardon, Sean F and Bischoff, Kendra (2011). Income Inequality and Income Segregation. *American Journal of Sociology* 116(4), 1092–1153.
- Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2019). Ipums usa: Version 9.0, american community survey, 2006. *Minneapolis, MN: IPUMS 9*, D010.
- Snipp, C Matthew and Cheung, Sin Yi (2016). Changes in Racial and Gender Inequality Since 1970. *The ANNALS of the American Academy of Political and Social Science* 663(1), 80–98.

## A Appendix Tables

**Table A1: Table of Matches**

	Hispanic and non-Hispanic white Count	Percent of Row Above	Asian and non-Hispanic white Count	Percent of Row Above
Total Observations in 2005 ACS	1,319,000		1,295,000	
Matched to PIKS	1,196,000	0.91	1,193,000	0.92
Merged to W2 Data in 2005	988,000	0.83	985,000	0.83

Source: ACS 2005 and IRS 2005–2014 W-2 and 1099 data. Note: Numbers have been rounded to comply with the Census Bureau’s disclosure-avoidance guidelines. Data approved for dissemination by CBDRB-FY19-370.