Heterogeneity and Intergenerational Mobility

Jonathan Rothbaum*

October 29, 2015

DRAFT – DO NOT DISTRIBUTE

Abstract

I use data from Census surveys linked to administrative tax records to explore heterogeneity in intergenerational mobility along two dimensions: 1) geography and 2) the relationship between parent characteristics and child earnings. First, I show that parent education and child race are highly predictive of child earnings even conditioning on parent earnings, and controlling for spatial variation in mobility does not affect these relationships. These characteristics are also highly correlated spatially with variation in mobility. I show that for children of families with incomes below the median, up to 40% of the spatial variation in mobility and up to 45% of the variation in causal impacts of mobility are due to parent sorting by race, education, and family type. Second, I use conditional and unconditional quantile regressions to explore heterogeneity in the relationship between these family and demographic characteristics and child earnings.

* Social and Economic Housing Statistics Division, U.S. Census Bureau. Email: jonathan.l.rothbaum@census.gov

I thank Raj Chetty and the participants of the Federal Reserve Community Development Research Conference on Economic Mobility, the Census Research Lunch, and the Census Labor Economics Reading Group for helpful comments and discussion.

1 Introduction

Understanding what characteristics of individuals or local areas are responsible for the variation in upward mobility has enormous implications for understanding intergenerational mobility and fostering equality of opportunity. In this paper, I show that parent education and child race are highly predictive of child earnings rank even after controlling for parent earnings and that geographic variation in these characteristics are also highly correlated with spatial variation in mobility. However, these relationships, such as the higher earnings of children of more educated parents and the lower earnings of Black sons, are not affected by controlling for the places in which they live.

Sorting, or spatial variation in a small set of family and demographic characteristics, can explain an important share of the spatial variation in mobility. For children of families with incomes below the median, up to 40% of the spatial variation found by Chetty, Hendren, Kline, and Saez (2014, hereafter CHKS) and up to 45% of the variation in causal impacts of place on mobility found by Chetty and Hendren (2015) are due to spatial variation in these family and demographic characteristics. Controlling for sorting disproportionately affects the South. Nearly all of the areas whose relative performance improves after adjusting for sorting are in the South.

I further explore heterogeneity in intergenerational mobility using conditional and unconditional quantile regressions, which allow me to investigate two separate but related issues. First, how do changes in a given parent characteristic relate to the distribution of child outcomes conditional on other characteristics? Second, how do these changes relate to the unconditional distribution of child outcomes? For example, I show that conditional on race, parent education, family type, and parent earnings, median and high achieving Hispanic children earn more than Whites, but that this relationship is entirely concentrated in otherwise low earning Hispanic children. In other words, these high achieving Hispanic children, in a conditional quantile regression, come from families that otherwise predict low child earnings, which is reflected in the unconditional quantile regression.

The high data requirements for studying intergenerational mobility have limited researchers' ability to explore heterogeneity. The results in this paper are made possible by the data used: administrative earnings data linked to two Census surveys, 1) the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and 2) the Survey of Income and

Program Participation (SIPP). The surveys provide the connection between parents and children and information on the characteristics of the parent families. The administrative data provides longitudinal earnings from W-2 records from 1978-2012. While the resulting sample is much smaller than the one used by CHKS and Chetty and Hendren, it is also much larger than the samples in other longitudinal surveys.[1]

Geographic variation in intergenerational mobility has been well documented in the economics and sociology literatures. Comparing mobility across countries, Jäntti et al. (2006) find that the United States has less intergenerational mobility than the United Kingdom and much less than the Nordic countries.[2] Corak (2013) shows that the United States, United Kingdom, and Italy have less intergenerational mobility than Canada, Germany, and the Scandinavian countries, among others.

Studies have also shown considerable geographic variation *within* the United States. Hertz (2008) found that mobility was lowest in the South and highest in the West Census regions. Using administrative data for nearly every child born in the United States born after 1980, CHKS are able to characterize the variation at a much finer geographic level. They find that parent income is more strongly associated with child income in the South and parts of the Midwest than in the Great Plains or Western states. However, even within regions, there is considerable heterogeneity. In the South and Midwest respectively, Texas and Pennsylvania are characterized by high upward mobility, while Georgia and Ohio are not.

As a first step in the analysis of potential causes of this heterogeneity, CHKS test for the relationship between a variety of local area characteristics and mobility. They find that certain factors are more strongly associated with differences in mobility, including spatial mismatch in access to jobs, inequality (GINI of the bottom 99%), school quality (measured by the high school dropout rate), social capital, and, especially, the fraction of single mothers.

Unfortunately, while CHKS have income data for nearly the universe of parents and children in the United States, they have almost no information on other characteristics of the parent families. As a result, they cannot easily control for these characteristics in their regressions and cannot distinguish between variation due to parent household sorting and the causal impact of location. To address the limitation, Chetty and Hendren (2015) study movers,

[1] The SIPP linked to administrative data has also been used to study intergenerational mobility by Mazumder (2005; 2011) and Gottschalk and Stinson (2014).

[2] Denmark, Finland, Norway, and Sweden

including controlling for parent fixed effects by comparing siblings of different ages at the time of the move, to establish the causal effect of place on child outcomes. They estimate that 50-70% of the observable variation across CZs reflects the causal effect of place.

In this paper, I take a different approach to study sorting and geographic variation in mobility. With data on parent household characteristics and parent and child earnings, I seek to answer several questions. First, does the observed spatial variation in mobility change our understanding of how parent characteristics relate to child outcomes? For example, Blacks are more likely to live in the South, and both CHKS and Chetty and Hendren show that upward mobility is lower in the South than in other regions. To what extent are worse outcomes for Black children caused by the places in which they grew up? Second, how does controlling for parent characteristics affect the relationship between place and child outcomes? By taking an alternative approach to control for parental sorting, this can also help validate the Chetty and Hendren causal results. This is potentially important because, even with their extremely large sample of movers (over 1.8 million children), their results suffer from high signal to noise ratios. In their raw estimates, few commuting zones (CZ) have statistically significant impacts on mobility.[3]

There is also a large literature on intergenerational mobility using long-term longitudinal surveys, which contain information on parent and child incomes and a very rich set of parent family characteristics, including the Panel Study of Income Dynamics (PSID) and National Longitudinal Surveys of Youth (NLSY). Intergenerational mobility has been studied using these surveys by many authors (Solon 1992; Zimmerman 1992; Hertz 2005; Mayer and Lopoo 2005; Bratsberg et al. 2007; Mazumder 2007; Lee and Solon 2009; Bhattacharya and Mazumder 2011, to name just a few). However, these data sets are orders of magnitude smaller than the administrative tax data used by CHKS and Chetty and Hendren.[4] This limits the types of

[3] Chetty and Hendren's have two preferred estimates for income mobility, parent family to child family income (at 26) for: 1) parent families below the median and 2) parent families above the median. For below median income families, only 12% of the 595 estimated CZs (comprising 10% of the all 709 CZs with estimates from CHKS) have statistically significant impacts at the 5% level. For above median families, only 8% of the 595 estimated CZs (comprising 7% of the all 709 CZs with estimates from CHKS) have statistically significant impacts at the 5% level. Chetty and Hendren use forecasting techniques to provide more precise estimates of the causal effects of place on child outcomes to address this signal-to-noise issue, but this precision comes from data on non-movers, whose parent characteristics cannot be controlled for in the same way.

[4] For example, referring to the PSID and NLSY papers cited above: Solon (1992): PSID 428, Zimmerman (1992): NLSY 826, Hertz (2005): PSID 6,273, Lee and Solon: (2009) PSID < 1,000 for non-pooled samples (single observation per child), Bratsberg et al. (2007): NLSY 1,999, Mazumder (2007): NLSY 4,944, Mayer and Lopoo (2005): PSID 1,567, and Bhattacharya and Mazumder (2011): NLSY 2,766.

questions that can be answered using the PSID or NLSYs. They are also more likely to be affected by measurement error than administrative data (Bound, Brown, and Mathiowetz 2001; Gottschalk and Huynh 2010).

The literature is rich in analyzing the relationship between parent characteristics and child outcomes. This includes work on the outcomes of siblings, twins, adoptees, natural experiments, etc., which is well summarized by Black and Devereux (2011). There is also work specifically focused on race on mobility. Hertz (2005; 2008), Bhattacharya and Mazumder (2011), and Mazumder (2011) show that Blacks experience less upward mobility and greater persistence of income across generations than Whites. Battacharya and Mazumder also find that differences Armed Forces Qualification Test (AFQT) results in the NLSY sample account for much of the gap in intergenerational mobility between Blacks and Whites.[5] Eide and Showalter (1999) investigate heterogeneity in intergenerational mobility using conditional quantile regressions in the PSID. They find that log of parent earnings (and income) and education is more associated with the log of child income at the bottom of the conditional distribution. I contribute to this literature by expanding on their work and documenting in more detail the considerable heterogeneity in mobility using quantile regressions and show how this heterogeneity is related to parent characteristics.

The paper is organized as follows. In section 2, I describe the data. In section 3, I estimate intergenerational earnings mobility and discuss how my results relate to previous work. In section 4, I discuss my results on family characteristics and geographic variation in intergenerational mobility. In section 5, I report my quantile regression results and discuss how heterogeneity and parent characteristics are associated with intergenerational mobility. Section 6 concludes.

2 Data

This paper uses survey data to construct a large sample of parents and children linked with administrative longitudinal earnings data. The parent-child links and information on parent family characteristics come from two surveys conducted by the US Census Bureau, the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the Survey of Income and Program Participation (SIPP). The earnings data comes from W-2 earnings records

[5] The AFQT measures cognitive skills and was administered to NLSY participants during adolescence.

filed by employers with the Social Security Administration and the Internal Revenue Service and shared with the Census Bureau in the Detailed Earnings Record (DER) extract from the SSA Master Earnings File. Individuals are linked between the Census surveys and the DER by matching survey respondents to their Social Security Numbers (SSN).[6] Prior to the construction of the Census survey-administrative data set, the SSNs are removed from the data and individuals are given a Personal Identification Key (PIK) to enable the linkage between data sets.

The CPS ASEC data available for linking to the DER and used in this paper comes from the following survey years: 1991, 1994, 1996-2009. The SIPP data used in this study comes from an internal data product at the US Census Bureau, the SIPP Gold Standard File (GSF). I use data from SIPP survey years 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008 panels. In the CPS ASEC only children aged 15 and older were given a PIK to allow matching to the DER, and I only include children observed in their parent household up to age 18 in my sample. The DER earnings file contains annual W-2 earnings information from 1978 to 2012.

To be included in the CPS-SIPP/DER (CSD) sample, each parent-child pair must be matched to their individual SSNs. A pair is successfully matched if the child and all parents (both parents in two-parent families and the individual parent in one-parent families) are successfully matched. The match rates for the CPS ASEC and SIPP samples by child age cohort is reported in Table 1. For all cohort groups in both surveys, the average match rate is 73%.

This data has a number of advantages over the data that has been used in the literature. First, in comparison to the administrative tax data used by CHKS and Chetty and Hendren, the CSD data contains a wealth of information on parent family characteristics, including race, education, occupation, industry, health insurance coverage, etc. There are, however, several disadvantages as well. The most obvious is that the CSD sample is orders of magnitude smaller than the administrative tax records. In my baseline sample, there are 49,725 parent-child pairs compared to 9,867,736 in CHKS. In addition, the DER file only contains information about wage and self-employment earnings taxable by Social Security, but not about other taxable income sources. Finally, the 1040 data used by CHKS and Chetty and Hendren also contains information on the marital status and income of the children's spouses. This information is not present in the DER.

[6] The process by which CPS ASEC and SIPP individuals are linked to the DER file is described in Wagner and Layne (2014).

This limits me to analyses of intergenerational earnings mobility between parent families and their individual children.

The CSD data set is also much larger than comparable survey data, such as the PSID and NLSYs. Another advantage of the CSD over longitudinal surveys is that the CSD earnings are from administrative data, which may make them less subject to measurement error. However, relative to the PSID and NLSY, the CSD data contains much less information about the parent and child households, especially for the children as adults.

As the CSD data comes from surveys using stratified random sampling, for all regressions and summary statistics, I use the CPS ASEC and SIPP sample weights normalized by survey and cohort age.[7] As I combine observations from two surveys over multiple years, the weights are adjusted by cohort and survey, which is discuss in greater detail in Appendix A.

An important step in analyzing mobility is to determine at what ages to measure parent and child earnings as a proxy for lifetime earnings in the intergenerational mobility comparison. For parents, I average family earnings over the 5 years when the older parent is 40-44 years old. This was chosen for two reasons. First, the literature on life-cycle bias in estimates of intergenerational mobility suggests measuring income around 40 (Haider and Solon 2006). Second, this choice allows me to better compare my results to CHKS and Chetty and Hendren as they use a 5-year average of parent income in their analysis.

For children, I follow CHKS in focusing on children around the age of 30. They show that there is little lifecycle bias in rank-rank income mobility by age 30 in child income. To test for lifecycle bias in the CSD sample, I plot the rank-rank slope of intergenerational earnings mobility with child earnings measured over two years starting from age 24 to 32, shown in Figure 1.[8] Panel A shows the effect of measuring earnings by age for the full sample. The general trend is similar to that in CHKS with increases at younger ages and potentially slight decreases at higher ages, but few of the differences are statistically significant. Panel B shows the trend for sons, which is increasing up to about 29 and flat above. The slight downward trend in Panel A is due to a decrease in the rank-rank slope of individual earnings for female children.

[7] For a discussion of the use of weights in OLS, see Solon, Haider, and Wooldridge (2013).

[8] I assign ranks to each parent family by comparing them to all parents in any year of the CPS ASEC and SIPP in the same age cohort as the older parent. For children, I assign an earnings rank by comparing each child to the cohort of all individuals in the same age cohort observed in any year of the CPS ASEC and SIPP, whether the individuals were observed as children or adults. In both cases, the comparison groups are much larger for each cohort than my sample of parents or children in the intergenerational mobility analysis. For more detail, see Appendix A.

I have chosen to use average child earnings at 29 and 30 for the baseline sample to maximize the sample size and to match the period used in CHKS, where income was measured starting at 29-32 years old depending on the child's age cohort.

In Table 2, I report summary statistics for family and demographic characteristics for my baseline sample (parents at age 40-44 for the oldest parent and children at 29-30) from the sample of all parent-child pairs in the Census surveys (columns 1-3) as well as for those parent-child pairs where both generations were successfully matched to their SSNs (columns 4-6). The family characteristics are taken from the first observation in a Census survey.

For the sample of all parent-child pairs, the share of Blacks and Hispanics in the CPS ASEC and SIPP are not statistically significantly different from one another. However, for parent education and share of single parent families, there are statistically significant differences.[9] For example, the SIPP has a higher share of parents with no high school degree and a lower share of parents with a college degree.

Also, comparing the matched parent-child pairs (where the parents and the child have PIKs that indicate a successful match to their SSNs) to the full set in each sample shows that there is some selection into matching. Blacks, Hispanics, and families with less educated parents are less likely to be matched than whites and families with more educated parents. Despite this selection, the matched sample is broadly representative of the underlying full set.[10]

More details about the data, weighting, and ranking of parents and children is available in Appendix A.

3 Intergenerational Mobility of Earnings Rank

As a first step in the analysis of intergenerational mobility of earnings rank, I replicate to the extent possible the results in CHKS. In doing so, I show that the geographic variation they find is present in the CSD data. By replicating their results, I can build on their work using the

[9] I measure parent education as the level achieved by the highest educated parent. The differences in parent education between the CPS ASEC and SIPP are likely due to the differences in single parenthood rates as the highest parent level of education is likely higher in two-parent families than single-parent families (for example, if the "unobserved" parent in the SIPP is more educated) .

[10] I estimate children living in families with unmarried partners using a modified Persons of Opposite Sex Sharing Living Quarters (POSSLQ) method, as cohabiting partners was not an option on either survey in the years studied. I classify two adults as partners if in the household: 1) a child of one adult is present, 2) there are only two opposite sex, unrelated adults, and 3) the potential partner is at least 15 years older than the child. All results are robust to excluding potential partner matches from the sample.

additional information available in the Census surveys with more confidence that any findings are due to the additional information and not other differences in the income and earnings data or parent-child samples.

As noted previously, an important difference between the two data sets is that CHKS and Chetty and Hendren use data on a larger variety of income sources.[11] In the CSD sample, I have only wage and self-employment earnings from W-2 filings. Another important difference is that CHKS and Chetty and Hendren can infer the marital status from 1040 filings in each year for both parents and children. In the W-2 data available in the DER, that is not possible. The only data available on marital status and family composition in the CSD sample is for parent families in the year they are observed in the survey, which may not correspond to the year of earnings used in the analysis.

3.1 Rank-Rank Mobility

CHKS estimate the rank-rank mobility for a variety of definitions of income, including parent family income →child family income, parent family income →child individual earnings, etc. using the basic regression model

$$y_i = \alpha + \beta x_i + e_i \tag{3.1}$$

where y_i is the income/earnings rank for the child and x_i is the rank for the parents in parent-child pair i. They do not report any coefficient for mobility of parent family earnings→child individual earnings, which corresponds to what can be estimated using the CSD data. Therefore, I compare my results to the closest analogue in CHKS, parent family income→child individual earnings.

Measuring mobility from a rank-rank regression has a number of advantages. First, the relationship is linear, which CHKS show is not true for log income. Second, the inclusion and treatment of zeroes is straightforward, whereas with log income and intergenerational elasticity, the coefficient is highly sensitive to these decisions. Figure 2 shows a binned scatter plot of average child and parent rank from CHKS and the CSD sample. The linear relationship between parent rank and average child rank holds in both data sets. The CHKS slope is steeper than in the CSD earnings data, which is reflected in the regression coefficients as well.

[11] In their papers, income includes wage and self-employed earnings, taxable capital and property income, tax-exempt interest income, and as well as some government transfer income (Social Security, disability benefits, and unemployment benefits).

In Table 3, I report the rank-rank regression coefficients estimated by CHKS and using the baseline CSD sample with cohort weights, including for male and female children separately. The CHKS coefficient for parent family income→child individual earnings is 0.282 compared to 0.251 for parent family earnings→individual child earnings in the CSD sample. In both datasets, the coefficient of rank-rank mobility for male children is higher and for female children is lower than in the combined samples.

CHKS and Chetty and Hendren divide the United States into 741 commuting zones (CZ).[12] For each CZ (c) with at least 250 children in their sample, CHKS estimated the slope (α_c) and intercept (β_c) of the parent-child rank-rank mobility regression using their baseline sample and income definitions. Using this information for each parent-child pair i, I created a predicted child rank ($\hat{y}_{ic}$) based on the parent earnings rank (x_{ic}) and the parent commuting zone (at the time of first survey observation) so that:

$$\hat{y}_{ic} = \alpha_c + \beta_c x_{ic}. \tag{3.2}$$

This predicted child rank accounts for both the parent earnings and the spatial variation in mobility. I regressed the child earnings rank on the CHKS predicted rank, with the slope coefficients reported in Column (3) of Table 3.

3.2 Geographic Heterogeneity

From this analysis, it is clear that there are differences between the rank-rank slope found in CHKS and the CSD data, potentially due to the difference between ranks based on income and earnings and the difference in child family income (in CHKS) vs. individual income (in the CSD sample). However, the main result of the CHKS and Chetty and Hendren paper is the geographic variation in mobility. To show that this heterogeneity is also present in the CSD data, I estimate the correlation between the CHKS CZ-level rank-rank regression coefficients and with the CSD data using the regression:

$$y_{ic} = \alpha_c + \beta_c x_{ic} + e_{ic}. \tag{3.3}$$

However, given the CSD sample of 49,725 parent-child pairs across 573 CZs, there are only 87 children per CZ on average. CHKS only includes CZs in their analysis with at least 250

[12] The commuting zones were constructed by Tolbert and Sizer (1996) based on commuting patterns in the 1990 census. They analyzed 1990 decennial census data on county to county commuting patterns, identified those with strong commuting ties, and grouped them into 741 CZs. Unlike Metropolitan Statistical Areas (MSA), CZs cover the entire country (not including territories).

observations. In Table 4 Panel A, I vary the minimum number of CSD observations required in a CZ for it to be included in the analysis, and calculate the correlation between the CHKS and CSD CZ-level intercept and slope for the included areas. At one extreme, including only CZs with at least 1,000 observations leaves only five in the analysis and unweighted correlation coefficients of 0.97 and 0.92 for the intercept and slope respectively. At the other extreme, including CZs with three observations leaves 544 and unweighted correlations of 0.17 and 0.09 for the intercept and slope. At the CHKS minimum of 250 observations, there are 39 CZs in the CSD sample with unweighted correlations of 0.66 and 0.61 for the intercept and slope.[13]

While the correlations for CZs with 250 or more observations suggest that the spatial heterogeneity found in CHKS is present in the survey-linked data, the small number of CZs leaves room for doubt. It would be preferable to include more CZs in the analysis while avoiding the statistical noise from small samples that is apparent in Panel A.

To do so, I create CZ groups that combine areas with similar levels of mobility so that I can more precisely estimate the coefficients using the CSD data. I order the 709 CZs from most to least mobile by their rank-rank slope coefficient from CHKS. The CZs are then divided into k quantile groups. For example with $k = 50$, the first group contains the 14 CZs with the lowest slopes, the second group contains the 14 with the next lowest slopes, etc. By decreasing the number of quantile groups, I can increase the minimum number of parent-child pairs in each group to get more precise coefficient estimates at the cost of combining CZs together with a wider range of slopes and intercepts. I vary k from 5 to 50 in the analysis.

For each group, I estimate the benchmark CHKS slope and intercept by averaging the CHKS CZ values across the individual CSD observations.[14] I calculate the correlation between the CHKS and CSD slope and intercept terms respectively, shown in Table 4 Panel B. The standard errors were estimated using a bootstrap with 100 replications of this entire process from the initial CSD sample. With five CZ quantile groups (and about 142 CZs per group), the smallest group has 6,641 observations and the correlation between both the slope and intercept in the CHKS and CSD is almost perfect (0.98 for both). At 25 groups, there are at least 612

[13] In each case, weighting the correlation by number of observations in the CSD sample yields an equal or greater point estimate, but the general pattern is the same.

[14] For example, if there are three equally-weighted individuals in the CSD samples in a given group and they live in CZs with slopes of 0.20, 0.21, and 0.28 and intercepts of 0.39, 0.36, and 0.30. The CHKS benchmark slope would be $\frac{0.20+0.21+0.28}{3} = 0.23$, and the intercept would be $\frac{0.39+0.36+0.30}{3} = 0.35$.

observations in each, and the correlations are 0.86 for the intercept and 0.77 for the slope. At 50 groups, there are at least 215 observations in each group and the correlations are 0.73 for the intercept and 0.58 for the slope.

Taken together, the results in Table 4 indicate that there is very high correlation between the spatial heterogeneity found by CHKS and in the CSD sample. In other words, low (high) mobility CZs in CHKS are also likely to be low (high) mobility CZs in the CSD data.

3.3 Family and Demographic Characteristics and Intergenerational Mobility

In this section, I take advantage of a small set of the rich family and demographic characteristics available in the Census surveys. First, I use a simple semiparametric model with dummies to estimate the relationship between child rank and parent earnings decile (D_{di}, $d \in (1,2,\ldots,10)$) and a variety of parent family and child demographic characteristics (h_i)

$$y_i = \sum_{d=2}^{10} (\beta_d D_{di} + \beta_{dh} D_{di} h_i) + \beta_h h_i + e_i \qquad (3.4)$$

where h_i includes child gender, race, and Hispanic status, the education level of the most educated parent (in 4 categories: 1) < high school, 2) high school, 3) some college, and 4) college or graduate), unmarried cohabiting partners, single parent, and whether the older parent was a teen parent. The expected child rank by race and decile is plotted in Appendix Figure 2.[15]

The results suggest that the relationship between parent rank and child rank is linear for each race, although the slopes may differ. There is no clear evidence for nonlinearity in child outcomes on other parent characteristics from the regressions either (not shown). As a result, I proceed using a simpler parametric model of family and demographic characteristics and parent and child outcomes:

$$y_i = \beta_x x_i + \beta_h h_i + \beta_{xh} x_i h_i + e_i. \qquad (3.5)$$

In Appendix Table 1, I test versions of this model with some β's set to zero to select a baseline model to proceed. Each model uses weighted OLS with the cohort weights discussed in Appendix A, although the results are largely unaffected by the use of weights. The baseline model in column (3) includes the interactions of Black, Hispanic, and female with parent rank

[15] The expected values are for children of high school graduates, married couples, and non-teen parents, which was chosen because each is the most common value for the respective category of dummies. A more detailed set of interactions between race and education groups at each decile was tested but the standard errors were too wide for the regressions to be informative given the small samples in many race-parent education-parent earnings decile cells.

along with interactions between parent rank and less than high school education and college or better education. I have left in these two parent education interaction terms as each is significant at the 10% level in one of the weighted and unweighted regressions in column (2), and for both, the point estimates are large in magnitude.

In the baseline model (also shown in Table 5, Column (1)), there are a number of notable results. The coefficients on Black, both the dummy (intercept) and dummy interacted with rank (slope), are statistically significantly different than for Whites. The greater slope for Black means that parent earnings are a more important determinant of outcomes for Black sons. The lower intercept means that Black sons start at a considerable disadvantage before accounting for parent earnings. The magnitude of the difference is striking. Given a White son whose parents' earnings are at the median, for a Black son to have the same expected rank, his parents would have to be at the 99th percentile. For a White son with parents at the 25th percentile, a Black son's parents would need to be above the 80st percentile for their children to have the same expected rank. Black females on the other hand are not disadvantaged relative to White females. They start with nearly identical intercepts (36.2 for blacks and 36.7 for whites, not statistically significant), and given the greater slope for Blacks, the expected rank increases faster in parent earnings for Black females than for Whites.

Another striking result is just how important parent education level is for child earnings even controlling for parent earnings. The expected child earnings rank is the same for a child with a parent with a college degree and another with a parent with a high school diploma whose income is more than 39 percentile higher in the parent rank distribution ($\frac{8.18}{0.209} = 39.2$). By the same token, the expected rank is the same for a child with a high school educated parent and another with no parent that completed high school and a 32 percentile higher parent earnings rank ($\frac{6.74}{0.209} = 32.3$). In both cases, the point estimates on the interaction terms suggest this gap narrows at higher parent income levels. Despite the very large effects of parent education, the coefficient on parent rank is largely unchanged from the earnings-only regression, as it declines from 0.251 to 0.209, or 17%. Parent education is an important predictor for child earnings and not only, or even primarily, because more educated parents earn more than less educated ones.

4 Results Controlling for Spatial Heterogeneity

4.1 Baseline Model and CHKS Predicted Ranks

However, if spatial heterogeneity is correlated with parent characteristics or child race, then the regression coefficients in the baseline model could be biased. For example, if Black sons predominantly live in low mobility CZs, their low upward mobility could be due to their location and not other unobserved characteristics correlated with their race.

This bias is suggested by CHKS in their Figure IX. As their individual-level data does not contain information on race, they analyze race and upward mobility by restricting their analysis to zip codes that are greater than some fraction White. As they increase the fraction of Whites in the zip code necessary for inclusion (and implicitly remove many Black children from the analysis), they find very little change in the upward mobility observed at the CZ level.

Next, I show the extent to which these parent and demographic characteristics are correlated with upward mobility in the CSD data. For each CZ quantile group used in Table 4, I calculate the correlation between the average prevalence of each characteristic and the observation-weighted average of the CHKS intercept and slope term from the CZ regression. I also calculate the CZ-level correlation between each characteristic and the CHKS intercept and slope using CZ-level aggregates calculated from 1990 Decennial Census longform microdata.

The results are shown in Table 6. Black children, high school educated parents, teen parents, and single parents live disproportionately in low mobility areas. College educated parents and, to a lesser extent, Hispanics live in high mobility areas. Therefore, the parent characteristics that are associated with worse (better) child outcomes are also concentrated in areas that are also associated with worse (better) child outcomes.

To test whether spatial heterogeneity is biasing the baseline results, I conduct two additional tests. In the first, I replace the parent rank and all interactions of rank in the baseline model with the CHKS CZ-based predicted child rank. The predicted rank accounts for both parent earnings and differences in mobility across CZs. Given the spatial heterogeneity present in the CHKS and CSD data, if CZ level differences in mobility were causing the lower upward mobility of Black sons, the coefficient on Black in the baseline predicted rank regression should be lower than in the parent rank regression.

In the second test, I include the gender-rank causal estimates from Chetty and Hendren. For each individual, this term is based on their gender and their parents' income relative to the median. This gender-rank causal estimate is included as an intercept term (as implied by Chetty and Hendren) and interacted with parent earnings as a slope term.[16] As in the first test, if the local area were responsible for the lesser or greater mobility of groups living predominantly in them, the coefficients should differ from the baseline regression.

The results are shown in Table 5. Column (1) shows the baseline model results for child rank regressed on parent rank, and column (2) shows the results using CZ-level predicted rank from CHKS in place of parent rank. For Black children, the coefficient on the dummy changes slightly in magnitude, but the difference is not statistically significant and the change is in the wrong direction. The point estimate for the intercept would indicate *less* upward mobility for blacks controlling for spatial heterogeneity, not more. The slope (black interacted with rank) does decline relative to the expected value (of about three times the slope for parent rank), but the difference is not statistically significant.[17]

What has changed is that the spatially adjusted slope coefficient for blacks is no longer statistically significantly different from whites, although it is still positive, implying an additional 0.03 percentile change in the expected black child rank for a one percentile increase in their parents' rank. However, while this result could be interpreted as evidence that blacks do not experience less relative mobility than whites conditional on spatial heterogeneity, it also means that black children of high income parents may not perform more like white children of similar parents (relative to children of each race from poor parents) as is indicated by the model in column (1). Instead, it implies that the better performance of black children from high earning parents may be because they are more likely to live in high mobility areas than black children from low earning parents.

[16] The gender-rank causal estimate for a given individual comes from four results from Chetty and Hendren: 1) causal estimates for below median males, 2) below median females, 3) above median males, and 4) above median females. Each child is assigned the CZ-level term that corresponds to that child's gender and parent income rank relative to the median. The results are nearly identical if the more general below/above median estimates that do not differentiate between genders are used.

[17] The expected change in CHKS predicted rank from a one percentile increase in parent rank is 0.341 (from the CHKS baseline regression). Therefore, a slope coefficient that was unaffected by the spatial heterogeneity would be scaled by $\frac{1}{0.341} = 2.93$ from model (1) to model (2). This can be seen in the parent rank term. The parent rank coefficient is -0.209 in model (1) and -0.602 in model (2) with a ratio of $\frac{-0.602}{-0.209} = 2.88$, very close to 0.293.

It is striking in comparing columns (1) and (2) just how little some of the coefficients on parent characteristics change. The level coefficients (uninteracted with parent rank) on child race, gender, teen parenthood, unmarried partners, and some college education are nearly identical in magnitude and none of them are statistically significantly different in the two models. The relationship between these characteristics and child rank is unaffected by the spatial heterogeneity in upward mobility. The other coefficients on parent education do change in magnitude, but not statistically significantly due to the large standard errors in model (2). However, the changes in the point estimates for the dummy and interacted variables would indicate an ever greater return for the child to parent education than in the baseline, especially for children from low income families. Therefore, it does not appear that by ignoring spatial heterogeneity in mobility, regression results for child earnings rank conditional on parent and family characteristics are biased upwards.

The same is true when including the Chetty and Hendren causal results in model (3) in addition to the variables in the baseline model. In other words, very little, if any, of the lower upward mobility experienced by Black sons is due to the lower mobility of the places in which they live. The same is true for the relationship between place and each of the other family and demographic characteristics tested in the model.

As an additional check on the relationship between place and child outcomes, I ran each model with five CZ-level covariates that CHKS show are correlated with mobility. These include: 1) segregation and spatial mismatch (fraction with a commute under 15 minutes), 2) income inequality (GINI for the bottom 99%), 3) school quality (high school dropout rate), 4) social capital (index constructed and used by Goetz and Rupasingha (2006)), and 5) rate of single parenthood (fraction of single mothers). As before, the results are unaffected by the inclusion of the covariates in any of the models.

4.2 Spatial Variation Adjusting for Family and Demographic Characteristics

Next, I attempt to calculate the share of the spatial variation found by CHKS and Chetty and Hendren that can be attributed to sorting. To do this, I calculate an adjusted rank for each child that controls for the observable family and demographic characteristics. This is possible because the regression coefficients are largely unaffected by controlling for the spatial variation in CHKS and causal estimates in Chetty and Hendren. I create an adjusted rank variable, $y_{i,adjusted}$, by subtracting the expected impact of each family and demographic characteristic from the observed child rank using the coefficients from the baseline model in Table 5 column (1), where

$$y_{i,adjusted} = y_i - \hat{\beta}_h h_i - \hat{\beta}_{xh} x_i h_i. \tag{4.1}$$

This adjustment normalizes individuals in each group to the model baseline group of White sons of married, high school educated parents. This was chosen as the baseline as it normalizes individuals to the most common group in each category in the CSD data. Figure 3 shows an example of how the adjustment affects CZ-level mobility estimates.

To test the impact of the adjustment, I again divide the individuals into k CZ quantile groups ($p = 1, \ldots, k$) based on the rank-rank slope in each CZ. I then calculate the adjusted slope and intercept term at the CZ-group level from the regression

$$y_{ip,adjusted} = \alpha_{p,adjusted} + \beta x_{ip} + e_{ip,adjusted}. \tag{4.2}$$

I also estimate an adjusted intercept and slope from the 1990 census data at the CZ-level. I assign each child a parent income rank based on their parents' position in the national income distribution of parents in the age cohort of the older parent. The rank is based on the total family income in 1989 reported in the survey. Using this imputed parent rank, I assign a predicted rank to each child using the CHKS CZ-level slope and intercept estimates. I then adjust each predicted child rank as in equation (4.1). Finally, I calculate a CZ-level adjusted slope and intercept using the regression equation (4.2) for each CZ.

To understand how much of the heterogeneity can be accounted for by the observable characteristics, I calculate a variety of measures for both unadjusted and adjusted intercepts ($\alpha_{p,adjusted}$ and α_p) and slopes ($\beta_{p,adjusted}$ and β_p). First, I regress the adjusted terms on the unadjusted ones as in: $\alpha_{p,adjusted} = \gamma_\alpha + \delta_\alpha \alpha_p$ and $\beta_{p,adjusted} = \gamma_\beta + \delta_\beta \beta_p$. I also estimate various measures of dispersion, including the variance, coefficient of variation, and mean absolute deviation for the adjusted and unadjusted intercept and slope. For each dispersion measure, I calculate the percent of the variation that is reduced by including the observable characteristics as $1 - \frac{\text{adjusted variation}}{\text{unadjusted variation}}$.

The results are shown in Table 7. For each of the estimated measures of variation accounted for by demographic characteristics, the 50 CZ group estimates and the 1990 census estimates are very close in magnitude. As the 1990 census estimates do not require grouping CZs together, I treat them as the preferred estimates. From the reduction in variance, the observable characteristics in the baseline model explain 50% of the variation in the intercept and 36% of the variation in the slope. This leaves 50% of the variation in the intercept and 64% of the variation in the slope unexplained. While the estimates are not directly comparable, this is very similar to the Chetty and Hendren estimate that 30%-50% of the variation in mobility is due to sorting. From the coefficient of variation results, it appears that the reduction in variance is being driven by a lower mean slope but less dispersion and not a lower mean in the intercept.

To more directly estimate the extent to which sorting can explain the spatial variation in mobility, I use the adjusted 1990 census estimates to create a counterfactual that holds the characteristics constant from the baseline model. Figure 4 shows CZ-level maps of CHKS expected child income rank for children with parent family income at the 25th and 75th percentile in Panels A and B respectively.[18] As previously discussed, in both maps there is less mobility for children from low income families (shown in red) in the South and parts of the Midwest.

In Figure 5, I show the magnitude of the parent characteristic adjustment by CZ, again for children with parent earnings at the 25th and 75th percentile. Areas in dark red, received a larger demographic and family characteristic adjustment, meaning that children in these CZs come from families with characteristics that are associated with lower earnings as adults. The South Mid-Atlantic, and parts of the Midwest are disproportionately populated by children who have lower expected earnings based on their family and demographic characteristics.

The maps in Figure 6 show the adjusted expected child rank for children from families at the 25th and 75th percentile. Although the maps in Figure 4 and Figure 6 are similar,[19] a few notable differences emerge. First, the South is no longer such an outlier in terms of its low upward mobility. Second, the West and areas of the Midwest, especially for children of high income families in the West and low income families in Midwest, also have worse expected outcomes in the adjusted maps than in the unadjusted ones. I also calculate the reduction in CZ-level variance in expected child rank at the 25th and 75th percentiles as a result of the adjustment. Sorting accounts for much more of the variation in outcomes of children of low income families than high income ones. I find that 40% of the variance in for children at the 25th percentile and 9% of the variance at the 75th percentile can be explained by the characteristics in my baseline

[18] Given the linear relationship between parent and child income, this is equivalent to showing the expected rank of all children from parent families with 1) below and 2) above median incomes respectively.

[19] The unweighted CZ-level correlation between the unadjusted and adjusted expected child rank from families at the 25th and 75th percentile of the income distribution is 0.93 and 0.97 respectively. The corresponding population weighted (by the 2000 population) correlations are 0.86 and 0.96.

model. Weighted by the 2000 population, the variance reduction is 45% of the variance at the 25th percentile and 6% at the 75th percentile.

Next, I compare how the adjusted CZ-level mobility results compare to those found by Chetty and Hendren. In their paper, Chetty and Hendren first calculate "raw" causal estimates of the impact of each CZ on mobility from a sample of 1,869,560 movers. They estimate this effect by analyzing movers from each pair of possible origin-destination CZs. Even with such a large sample, given the large number of potential CZ-to-CZ moves (and the small number of movers to and from CZs with smaller populations), they cannot estimate these raw causal impacts with high precision. In their raw estimates, Chetty and Hendren find that fewer than 10% of the 709 CZs (the set with predictions in CHKS) have a statistically significant impact on child outcomes (calculated separately for children from below and above median families). They estimate that 71% of the variation in their raw causal estimates is due to sampling variation as opposed to the causal effects of place.

To overcome this problem, they create a "forecast" prediction of the causal effect of place by combining the raw causal estimate with the predictions from non-movers. In practice, this forecast involves taking the weighted average of two terms: 1) the raw causal estimate and 2) the mobility experienced by permanent residents multiplied by a constant.[20] For CZs with more precise raw causal estimates, greater weight is given to the first term and for CZs with fewer movers and less prices raw estimates, more weight is given to the second term. Their preferred estimates are the forecasts.

For each causal estimate, the raw and forecast, Chetty and Hendren calculate estimates separately for children from above and below median households (alternatively denoted as estimates for children of families at the 25th and 75th percentiles of the income distribution). Figure 7 shows the map of the forecast estimates by CZ. Although it is very similar to the adjusted estimates in Figure 6, the South seems to suffer from lower mobility and the West and Southwest seem to have higher mobility according the Chetty and Hendren estimates than the family and demographic adjusted estimates.

Next, I explore why my results may differ from their preferred estimates. In Figure 8 and Figure 9, I show scatter plots of the raw and forecast estimates for the 25th and 75th percentile in

[20] The constant is the β coefficient from the regression of the raw causal impacts on the expected outcomes of permanents residents.

Panel A.[21] The correlation between the raw and forecast estimates is 0.49 for the 25th percentile children and 0.16 for 75th percentile children.

In Panels B-E of the two figures, I plot the relationships between each of the causal estimates and 1) the Chetty and Hendren estimates of mobility for permanent residents (non-movers), 2) the CHKS estimates and 3) the CHKS estimates adjusted for family and demographic characteristics. In each panel, both variables are normalized to have mean zero and standard deviation 1, so all regression coefficients are correlations (which is not the case in Panel A). For 25th percentile children, the forecast is highly correlated with the three potentially biased measures of mobility, 1) the prediction for non-movers from Chetty and Hendren, 2) the CHKS estimates, and 3) the adjusted estimates (about 0.9). Not surprisingly, given the lack of precision in the estimates, the correlation between the raw causal impact and these three measures is much lower (about 0.5). For 75th percentile children, the forecast is also highly correlated with the three estimates (about 0.8). However, there is very little correlation between the raw estimates and any of the three other predictions (about 0.15). In other words, there is no evidence that the raw causal estimates are more closely related to the permanent residents or unadjusted CHKS estimates than the ones adjusted for sorting by race, education, and family type. It is likely then that the difference between the forecast estimates and the adjusted ones are due to the inclusion of the permanent resident data in the forecast.

Therefore, I create an adjusted forecast that reduces the sorting bias. To do so, I replace the unadjusted Chetty and Hendren permanent resident data with the adjusted CHKS CZ-level estimates at the 25th and 75th percentile of the distribution. The difference between the adjusted forecast and the Chetty and Hendren forecast is shown in Figure 10. For below-median children, the causal mobility estimate is most improved in the South. For many CZs in the South (in dark red), the improvement is between 0.5 and 1.5 standard deviations of the original CZ distribution. For above-median children, the areas affected are more broadly distributed across the South, along the Great Lakes and in California. Figure 11 shows the forecast causal estimates with the family and demographic characteristic adjustment to non-mover mobility. Although the South is still generally characterized by low upward mobility for below median children, it is no longer such an outlier as it is in Figure 7.

[21] In all panels in these figures, the variables are standardized to have mean 0 and standard deviation 1 without population weights. The results are available upon request for the figures with 2000 population weights, but qualitatively they are the same.

Figure 12 shows a scatter plot of the Chetty and Hendren forecast causal estimates and the estimates with the family and demographic characteristic adjustment to permanent resident mobility. The two estimates are very highly correlated (0.89 at the 25th percentile and 0.92 at the 75th percentile). However, the high correlation is largely due to CZs outside of the South. At the 25th percentile, the South contains 93% of the 57 CZs where the adjusted causal estimate is at least ½ of a standard deviation greater than the Chetty and Hendren forecast.[22] The variance in the causal estimates is reduced by 45% for 25th percentile children and 18% for 75th percentile children. However, it should be noted that this reduction in variance is highly sensitive to which measure is used as the pre-adjustment baseline for permanent residents.[23]

5 Heterogeneity of Intergenerational Mobility

In the final section of the paper, I explore heterogeneity of a different sort using conditional (Koenker and Bassett, Jr. 1978) and unconditional (Firpo, Fortin, and Lemieux 2009) quantile regressions. I use the quantile regressions to investigate the relationship between the baseline parent characteristics and child outcomes. Conditional quantile regressions allow me to test whether these characteristics are associated with different levels of mobility for low vs. high achieving children conditional on observable characteristics. For example controlling for parent rank, education, and family type, do high-achieving Hispanic children have earnings or earnings ranks that are greater than high-achieving white children?

First, I regress child rank on parent rank using conditional quantile regressions and cohort weights for the CSD data. I also estimate the quantile regression slopes from CHKS using data they have made available on their website.[24] The results are shown in Appendix Figure 4. In both data sets, there is considerable heterogeneity in child outcomes by conditional quantile with

[22] In this case, the South is defined as the 11 states in the Confederacy: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia. These 53 CZs represent 23% of all CZs in the South. The four other sufficiently improved CZs represent 1% of those outside of the South. As defined by Census region, the South also includes Delaware, Kentucky, Maryland, Oklahoma, Washington, DC, and West Virginia. No CZs in these additional states have a positive causal adjustment of ½ of a standard deviation.

[23] The geographic distribution of the adjustment is not affected. In my baseline (shown in the figures), I use the CHKS estimates for permanent residents at 29-32 as these are used to create the characteristic adjustment. If I use the Chetty and Hendren estimates for permanent residents at 29 as the baseline, the variance is reduced by 31% at the 25th percentile and 17% at the 75th percentile, and all 48 CZs with a ½ standard deviation improvement at the 25th percentile were in the Confederacy. With the Chetty and Hendren estimates at for permanent residents at 26 as the baseline (as in their paper), the variance increases by 12% at the 25th percentile and declines by 9% at the 75th percentile and 95% of the CZs with a ½ standard deviation improvement at the 25th percentile were in the Confederacy. Those results are shown in Appendix Figure 3.

[24] http://www.equalityofopportunity.org

a nearly identical pattern. For children in low quantiles, or "low-achievers" conditional on parent earnings, there is a weaker relationship between parent rank and child rank. For children in high quantiles, or "high-achievers", there is similarly a much weaker relationship between parent rank and child rank. It is the "average" children (roughly the 25th to 75th percentile) whose rank is highly dependent on their parents' ranks. The slope near the median is almost 0.4 in the CSD data, nearly 60% higher than the OLS slope. In other words, it is for these average children that mobility of rank is more limited, but the subset of conditionally low and high earners seem to fall behind or excel mostly independently of their parents' position in the distribution.[25]

I also use unconditional quantile regressions to analyze the upward mobility. Unconditional quantile regressions were developed by Firpo, Fortin, and Lemieux (2009) to analyze how changing the marginal distribution of explanatory variables would affect the marginal distribution of some variable Y. In this case, the question is how would the sample distribution of child earnings (whether in ranks, dollars, or logs) be affected by changing the sample distribution of parent earnings or of the family and demographic characteristics?

For example, if conditional on the observed characteristics, high achieving Hispanics do outperform high achieving whites, where in the marginal or *unconditional* distribution is this effect present? Since the conditional quantile regression controls for the model characteristics, it is possible that the high-achieving Hispanics could be anywhere in the marginal distribution of child earnings. If Hispanic high achievers are primarily children of low-earning parents without a high school education, they may be concentrated at the bottom of the earnings distribution despite their relative success. However, if they are primarily children of high-earning college-educated parents, their additional earnings would be concentrated higher in the child distribution.

I test for the relationship between child earnings and parent earnings and other family characteristics in three separate sets of regressions. In each set, I regress a measure of child achievement on parent achievement and dummies for each of the family and demographic characteristics in the baseline model in Table 5 for both conditional and unconditional quantile regressions. In the first set of regressions (rank-rank), I regress individual child earnings rank on parent family earnings rank. In the second set (earnings-rank), I regress individual child earnings

[25] This inverted U-shape is potentially a mechanical result of the rank-rank specification. I explore alternative transformations of earnings in the family characteristic quantile regressions later in this section.

on parent family earnings rank. In the third set (log earnings-log earnings), I regress the log of individual child earnings on the log parent family earnings, with zeroes recoded to 1.

5.1 Rank-Rank Quantile Regressions

The results for the first set of regressions, with child rank regressed on parent rank, are shown in Figure 13 and Figure 14. The rank-rank results exhibit a U-shaped (or inverted U-shaped) pattern across most variables. This is driven in large part by the rank transformation (as will be shown in the other quantile regression specifications) as in the middle of the income distribution, smaller income changes lead to larger rank changes.

However, the U-shape is not present to the same degree in all variables. The coefficient on teen parents is more negative at higher quantiles than lower ones, which means that there is a larger gap between high-achieving children of teen parents and other high-achieving children than for low-achievers, conditional on the other characteristics. It is again striking how large the coefficients for Black males are. The median Black son, conditional on his family characteristics, ranks 20 percentage points lower than a similar White son. The coefficients for Black interacted with female generally more than offset the large negative coefficients on Black. At the 15th percentile, Black females have an expected rank that is 7.1 percentile points higher than Whites. The difference is greater than 6 percentile points up to the 45th percentile after which it declines steadily to 0.01 percentile points at the 60th percentile (not statistically different from zero) and become a statistically significant -2.5 at the 70th percentile, -3.5 at the 80th percentile, and -4.5 at the 90 percentile.

Also notable is that for Hispanics the results differ from the OLS. Hispanic children across the distribution outperform White children by about 2 percentage points (from the 25th to 85th quantiles). In the baseline OLS model, the point estimate for Hispanic corresponds to the conditional quantile regression results, but the OLS estimate is not statistically significant. Otherwise, the results mirror those in the OLS regressions; female children earn less than males, and parent education is an extremely important predictor of child outcomes even after conditioning on parent earnings. The effect of parent education is stronger for children above the 40th percentile in the conditional distribution for all parent education types (and below the 85th percentile). Lower parent education is more harmful and more parent education is more beneficial to middle and high achieving children.

The unconditional quantile regression results are shown in Figure 14. The adverse outcomes for Black males affect children throughout the distribution, but are especially concentrated from the 15^{th} to 60^{th} percentiles. For Black females, the positive effects are heavily concentrated in the bottom half of the earnings distribution. The sum of Black and Black interacted with female is 8.5 at the 20^{th} percentile, but declines to -2.9 at the 60^{th} percentile and is negative but not statistically different from zero at the 70^{th} percentile and above.

For Hispanic children, the greater upward mobility they experience in the conditional quantile regression primarily affects children who are in the bottom 40 percent of the marginal distribution. This suggests that the Hispanics that outperform Whites are primarily from families with characteristics that would likely place them in the bottom half of the child earnings distribution. The effect of parent education differs by type. Not surprisingly, children whose parents did not graduate high school are primarily affected at lower ranks in the child distribution, whereas children from college educated parents are concentrated at higher earnings.

5.2 Earnings-Rank Quantile Regressions

In the second set of quantile results, I regress child earnings (in 2012 dollars) on parent rank and the baseline characteristics. The conditional quantile regression results in Figure 15 show the extent to which the U-shaped results in the previous section were due to the rank transformation. The relationship between parent rank and child earnings increases monotonically across conditional quantiles for most variables. Up to the tenth percentile, there is no statistically significant relationship between parent rank and child earnings. The relationship is steeply increasing from the 15^{th} to 40^{th} percentile before leveling off somewhat. At the 15^{th} percentile, a 1 point increase in parent earnings rank results in a \$41 increase in child earnings. This increases to \$204 at the median, \$257 at the 75^{th} percentile, and \$337 at the 90^{th} percentile. Therefore. for a high achieving child (at the 90^{th} percentile), conditional on other characteristics, having parents at the top of the distribution rather than the bottom is associated with a nearly \$34,000 increase in earnings at 30.

The earnings of Black sons relative to similar Whites also declines as they move up the distribution. At the 25^{th} percentile, the gap is \$9,365, increasing to \$14,528 at the median, \$16,272 at the 75^{th} percentile, and \$19,699 at the 90^{th} percentile. As in the rank-rank regressions, similar Black daughters outperform Whites at lower percentiles (up to the 55^{th} percentile in this case) with a peak of $4,308 a the 35th percentile, but earn less at higher ones (85th and above). Hispanics also statistically significantly outperform similar Whites from the 35th to 80th percentile, by about $1,000-2,000 at all points in that range.

Although the impact of education is monotonically increasing by child quantile, having a parent with a college education is especially beneficially to high achieving children, reaching $21,301 at the 90th percentile (from $7,020 at the median). At lower child percentiles, the negative impact of having no parent who graduated high school and a parent who graduated college are of the opposite sign but similar in terms of economic significance (~$4,000-$7,000 from 35th to 55th percentile)

Looking at the unconditional quantile regression results in Figure 16, we can see that Hispanics have a strong positive impact on the child earnings distribution at relatively low ranks (20th-45th percentile). Again, the impact of parent education on the child earnings distribution also differs by type. Children of non-high school graduates particularly depress earnings at low ranks (20th-60th percentile) and children of college graduates particularly increase earnings at higher ranks. For other characteristics, the conditional and unconditional quantile regression results follow the same patterns, meaning that the conditional results are not concentrated in a particular section of the marginal distribution.

5.3 Log Earnings-Log Earnings Quantile Regressions

The last set of quantile regressions uses the log-log specification to show how the prior results relate to the distribution of intergenerational earnings elasticity. In these regressions, I include all observations, with zeroes recoded as $1 prior to the log transformation. This is less problematic in quantile regressions than in OLS as quantile regression results are not as sensitive to outliers. The results are shown in Figure 17 and Figure 18.[26]

The earnings elasticity monotonically decreases from the 20th percentile from very high levels (0.53 at the 20th percentile) to below 0.05 above the 65th percentile. A similar pattern (in absolute values) is present for Black males and females and all females. For Black sons, the coefficient crosses -1 between the 35th and 40th percentiles. The implication of a -1 coefficient is that white males earn 170% more than Black males conditional on other characteristics. The

[26] In these figures, the results are shown from the 20th percentile and above as the coefficients at lower percentiles are dominated by the effect of zeroes. The full results are shown in Appendix Figure 5 and Appendix Figure 6.

coefficient drops to -0.32 at the 90th percentile, for a difference of 37% between Whites and Blacks. The combined coefficients for Black females is positive up to the median (0.63 at the 25th percentile, meaning Black females earn 88% more than similar Whites) declining to -0.10 at the 90th percentiles for a 10% difference in earnings.

For children of parents without a high school degree, the elasticity also declines in absolute value as child quantile increases. However, the magnitudes are always very large in economic terms, from -0.29 at the median (34% lower earnings) to 0.17 at the 90th percentile (18% lower earnings). For children with a parent who graduated college, the coefficient remains constant over much of the distribution before increasing at the top, from 0.30 at them median (36% higher earnings) to 0.39 at the 95th percentile (48% higher earnings).

In the unconditional quantile regression, having parents with a college education is associated with a large percentage increase in child earnings across the distribution, but especially at the bottom and top of the distributions (25th percentile: 0.51 for a 66% increase, 50th percentile: 0.32 for a 38% increase, and 90th percentile: 0.40 for a 48% increase). Again, the contribution of Hispanics is concentrated at the bottom of the distribution.

6 Conclusion

There are many dimensions of heterogeneity in intergenerational mobility that are only beginning to be explored as new and better data becomes available. In this paper, I explore just few of those using data from Census surveys linked to administrative tax data.

I build upon the work by Chetty, Hendren, Klein, and Saez and Chetty and Hendren on spatial heterogeneity in the United States by controlling for family and demographic characteristics. I show the degree to which parent education and child race are predictive of child earnings rank, even after controlling for parent earnings. These characteristics are also highly correlated across space with upward mobility. Black children and parents with less education are concentrated in low mobility areas and parents with a college education are concentrated in high mobility areas. I show that neither the lower earnings of low mobility groups (Black sons, children of less educated parents) nor the higher earnings of high mobility groups (children of college educated parents) are due to the places in which they grew up.

My results echo Chetty and Hendren's in showing that up to half of the spatial variation in mobility in the United States is due to sorting. However, controlling for this sorting by adjusting mobility for a small set of demographic and family characteristics effects how different areas rank in terms of upward mobility. Low mobility characteristics are heavily concentrated in the South and mid-Atlantic. After adjusting for these characteristics, the South is no longer such an outlier in terms of the low upward mobility it causes, although the South remains the region with the worst outcomes for below median children. I also adjust Chetty and Hendren's preferred causal estimates for sorting. Again, the South is particularly affected by this adjustment as nearly all of the CZs with a large improvement in their adjusted causal estimates are in the South. Although it is very sensitive to how the adjustment is done, I estimate that up to 45% of the variation in Chetty and Hendren's causal estimates for children of low income parents is due to sorting as well.

Finally, I explore the heterogeneity in the relationship between parent characteristic and child earnings using conditional and unconditional quantile regressions. This allows me to characterize 1) how the relationship between mobility and parent and family characteristics differs for low and high achieving children, such as how Hispanics earn more than Whites conditional on other characteristics, and 2) how these characteristics affect the marginal distribution of child outcomes. For example, Hispanics outperform similar Whites, but the effect is concentrated entirely in children with low earnings as adults.

References

Bhattacharya, Debopam, and Bhashkar Mazumder. 2011. "A Nonparametric Analysis of Black-White Differences in Intergenerational Income Mobility in the United States." *Quantitative Economics* 2: 335–79.

Black, Sandra E., and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics* 4 (11). Elsevier: 1487–1541.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. "Why the Apple Doesn't Fall Far: Understanding the Intergenerational Transmission of Human Capital." *American Economic Review* 95 (1): 437–49.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." *Handbook of Econometrics* 5. Elsevier: 3705–3843.

Bratsberg, Bernt, Knut Røed, Oddbjørn Raaum, Robin Naylor, Markus Jäntti, Tor Eriksson, Eva Osterbacka, and Eva Österbacka. 2007. "Nonlinearities in Intergenerational Earnings Mobility: Consequences for Cross-Country Comparisons." *The Economic Journal* 117 (519): C72–92.

Chetty, Raj and Nathaniel Hendren. 2015. "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates." *Harvard University Working Paper*.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129 (4): 1553-1623.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27 (3): 79–102.

Eide, Eric R., and Mark H. Showalter. 1999. "Factors affecting the transmission of earnings across generations: A quantile regression approach. " *Journal of Human Resources*: 253-267.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica* 77 (3): 953–73.

Goetz, Stephan J. and Anil Rupasingha. 2006. "Wal-Mart and Social Capital," *American Journal of Agricultural Economics*, 88 (5): 1304-1310.

Gottschalk, Peter, and Minh Huynh. 2010. "Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error." *Review of Economics and Statistics* 92 (2): 302–15.

Gottschalk, Peter, and Martha Stinson. 2014. "Is There an Advantage to Working? The Relationship between Maternal Employment and Intergenerational Mobility." *Survey of Income and Program Participation Working Paper*.

Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96 (4): 1308–20.

Hertz, Tom. 2005. "Rags, Riches, and Race: The Intergenerational Economic Mobility of Black and White Families in the United States." In *Unequal Chances: Family Background and Economic Success*, edited by Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves, 338–58. Princeton, NJ: Russell Sage Foundation and Princeton University Press.

———. 2008. "A Group-Specific Measure of Intergenerational Persistence." *Economics Letters* 100 (3): 415–17.

Jäntti, Markus, Bernt Bratsberg, Knut Røed, Oddbjørn Raaum, Robin Naylor, Eva Osterbacka, Anders Bjorklund, and Tor Eriksson. 2006. *American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States*. IZA Discussion Paper No. 1938.

Koenker, Roger, and Gilbert Bassett, Jr. 1978. "Regression Quantiles." *Econometrica: Journal of the Econometric Society* 46 (1): 33–50.

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91 (4): 766–72.

Mayer, Susan E., and Leonard M. Lopoo. 2005. "Has the Intergenerational Transmission of Economic Status Changed?" *Journal of Human Resources* 40 (1): 169–85.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *Review of Economics and Statistics* 87 (2): 235–55.

———. 2007. "Sibling Similarities and Economic Inequality in the US." *Journal of Population Economics* 21 (3): 685–701.

———. 2011. "Black-White Differences in Intergenerational Economic Mobility in the US." *Federal Reserve Bank of Chicago Working Paper 2011-10*.

Oreopoulos, Philip, Marianne E. Page, and Ann Huff Stevens. 2006. "The Intergenerational Effects of Compulsory Schooling." *Journal of Labor Economics* 24 (4): 729–60.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *The American Economic Review* 82 (3): 393–408.

Solon, Gary, Steven J. Haider, and Jeffrey Wooldridge. 2013. "What Are We Weighting For?" *National Bureau of Economic Research Working Paper 18859*.

Tolbert, Charles M., and Molly Sizer. 1996. *US Commuting Zones and Labor Market Areas: A 1990 Update*. *US Department of Agriculture Economic Research Service Staff Paper #9614*.

Wagner, Deborah, and Mary Layne. 2014. *The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software. CARRA Working Paper Series #2014-01.*

Zimmerman, David J. 1992. "Regression toward Mediocrity in Economic Stature." *The American Economic Review* 82 (3): 409–29.

Table 1: Match Rate Between Parent-Child Pairs in Surveys to Administrative Records

Child Cohort (By Birth Year)	Match Rate (%)		
	CPS ASEC	**SIPP**	**Total**
1972		70.4%	70.4%
1973	77.4%	73.2%	75.4%
1974	77.5%	73.7%	75.6%
1975	77.9%	77.0%	77.4%
1976	71.9%	77.3%	74.4%
1977	67.4%	77.2%	73.4%
1978	71.7%	75.9%	74.1%
1979	70.0%	77.5%	73.7%
1980	68.5%	77.7%	73.6%
1981	65.0%	78.6%	71.7%
1982	63.1%	77.7%	69.4%
Total	69.2%	76.9%	73.2%

This table show the match rate between parent-child pairs in the CPS ASEC and SIPP. Each parent-child pair is successfully matched if the child and all parents (both parents in two-parent families and the individual parent in one-parent families) are matched to a valid SSN. No observation weights are used to determine the match rate.

Table 2: Demographic Characteristics and Summary Statistics for Parent-Child Pairs

Variable	All Parent-Child Pairs (Matched and Unmatched)			Matched Parent-Child Pairs		
	CPS ASEC (1)	SIPP (2)	Total (3)	CPS ASEC (4)	SIPP (5)	Total (6)
Parent Family Earnings				66,318	65,804	66,049
Individual Child Earnings				32,856	35,368	34,172
Black	14.5%	14.3%	14.4%	12.9%	13.3%	13.1%
Hispanic	11.7%	11.0%	11.2%	9.2%	10.3%	9.8%
Highest Educated Parent						
< High School	11.9%	14.2%	13.2%	10.3%	12.8%	11.6%
High School	32.8%	33.7%	33.4%	32.6%	33.0%	32.8%
Some College	27.7%	28.2%	27.8%	28.6%	29.3%	29.0%
College or Graduate	27.7%	24.0%	25.5%	28.5%	24.9%	26.6%
Teen Parent	7.5%	7.8%	7.6%	7.1%	7.1%	7.1%
Single Parent	26.3%	33.9%	30.6%	25.2%	32.1%	28.8%
Partner Family	2.7%	2.5%	2.5%	2.4%	2.1%	2.3%
Observations	33,002	34,976	67,978	22,829	26,896	49,725

This table shows the DER earnings and family demographic information for the survey samples. The first three columns show the cohort weighted demographic information for children from all parent-child pairs in the CPS ASEC and SIPP samples. The second three columns show the cohort weighted demographic information for the parent-child pairs that were successfully matched to the their SSNs. The fraction of children from single parent families is one of the few variables that is statistically significantly different between the CPS ASEC and SIPP samples. The parent family earnings are the average when the older parent is 40-44 years old. The individual child earnings are the average when the child is 29-30 years old.

Table 3: Comparing CPS-SIPP/DER Earnings Mobility to CHKS

	CHKS Parent Family Income → Child Individual Earnings	CPS-SIPP/DER Parent Family Earnings → Child Individual Earnings	
		Parent Rank	CHKS Predicted Rank
	(1)	(2)	(3)
All Children	0.282	0.251	0.717
Males Only	0.313	0.296	0.842
Females Only	0.249	0.207	0.594

All of the coefficients are significant at the 1% level. CHKS do not report the slope of a rank-rank regression of child individual earnings on parent family earnings. The most comparable reported coefficient is of parent family income and child individual earnings (1). Column (3) is the regression of individual child earnings on the predicted child rank based on the parent earnings and the CHKS CZ-level mobility coefficients. The coefficient for all children is 0.717 which is nearly equal to the ratio of the CPS-SIPP/DER slope to the baseline CHKS slope $\left(\frac{0.251}{0.341} = 0.736\right)$. The regressions in columns (2) and (3) use the cohort weights discussed in the Online Appendix

Table 4: Spatial Heterogeneity in CHKS and the CPS-SIPP/DER

A. Correlation Between CZ-Level Rank-Rank Regression Coefficients

Min Observations in CSDfor Inclusion	# of CZs	Unweighted		CSD Observation Weighted	
		Intercept	Slope	Intercept	Slope
3	544	0.17	0.09	0.32	0.16
50	243	0.31	0.13	0.38	0.23
100	131	0.47	0.27	0.54	0.40
250	39	0.66	0.61	0.66	0.61
500	10	0.46	0.69	0.57	0.80
1000	5	0.97	0.92	0.97	0.94

B. Correlation of Rank-Rank Regression Coefficients for CZs Grouped by CHKS Slope

Quantile Groups	Intercept	Slope	Parent-Child Pairs in Smallest Quantile Group	Average Number of Parent-Child Pairs
5	0.98***	0.98***	6,641	9850
	(0.05)	(0.04)		
10	0.95***	0.94***	3,164	4925
	(0.09)	(0.07)		
20	0.94***	0.84***	789	2462
	(0.06)	(0.09)		
25	0.86***	0.77***	612	1970
	(0.12)	(0.10)		
50	0.73***	0.58***	215	985
	(0.01)	(0.004)		

Panel A reports the correlation between the slope and intercept of the CZ-level rank-rank mobility regression for the CHKS and CPS-SIPP/DER baseline samples using the regression equation for each CZ c of $y_{ic} = \alpha_c + \beta_c x_{ic} + e_{ic}$. Given the baseline CHKS sample size of 9,867,736 over 741 CZs, there are nearly 13,500 observations per CZ. The CPS-SIPP/DER sample contains 49,725 observations over 573 CZs, or about 87 individuals per CZ. CHKS limits their regressions to CZs with at least 250 parent-child pairs, yielding a sample of 709. Due to the small number of CZs with reasonably large samples for the rank-rank regression, I also calculate another measure of the correlation between the CHKS spatial heterogeneity and the CPS-SIPP/DER in Panel B. I divide the CZs into k quantile groups from lowest to highest rank-rank slope. For example, with 50 quantile groups, the first group contains the most mobile 14 CHKS CZs, and each quantile group contains on average 985 observations in the CPS-SIPP/DER. For each group, I calculated the intercept and slope in the CPS-SIPP/DER data and the observation weighted average of the CZ-level intercept and slope from CHKS. I then calculated the correlation between the two sets of coefficients. The standard errors were calculated using a bootstrap with 100 replications. The results are very similar with the CZs ordered into quantile groups by the intercept instead of by the slope (not shown).

Table 5: Intergenerational Mobility and Family Characteristics Controlling for Spatial Heterogeneity

Dependent Variable = Child Rank	Baseline			With CZ Characteristics		
Variable	**Parent Rank**	**CHKS Predicted Rank**	**CH Causal Prediction (with Rank Interaction)**	**Parent Rank**	**CHKS Predicted Rank**	**CH Causal Prediction (with Rank Interaction)**
	(1)	(2)	(3)	(4)	(5)	(6)
Parent Rank (CHKS Predicted in (2) and (5))	0.209***	0.602***	0.212***	0.206***	0.621***	0.203***
	(0.013)	(0.037)	(0.013)	(0.014)	(0.041)	(0.014)
Black	-15.64***	-17.45***	-14.62***	-15.86***	-20.56***	-15.32***
	(1.11)	(3.10)	(1.14)	(1.21)	(3.01)	(1.20)
Hispanic	1.66	-0.81	1.75	1.53	-2.70	0.84
	(1.49)	(5.69)	(1.48)	(1.36)	(5.55)	(1.33)
Female	-8.39***	-4.18*	-7.99***	-8.39***	-4.05*	-7.95***
	(0.83)	(2.35)	(0.74)	(0.82)	(2.28)	(0.74)
Black*Female	15.06***	15.05***	13.96***	15.04***	14.96***	13.66***
	(1.08)	(1.12)	(1.09)	(1.09)	(1.13)	(1.10)
Most Educated Parent						
< High School	-6.74***	-11.94***	-6.84***	-6.82***	-13.75***	-7.06***
	(1.03)	(3.78)	(1.05)	(1.02)	(3.65)	(1.04)
Some College	1.91***	1.98***	1.94***	1.99***	1.93***	1.95***
	(0.47)	(0.47)	(0.47)	(0.47)	(0.47)	(0.46)
College+	8.18***	13.60***	8.59***	8.37***	14.94***	8.44***
	(1.05)	(2.71)	(1.03)	(1.04)	(2.64)	(1.04)
Teen Parent	-3.160***	-3.227***	-3.191***	-3.061***	-3.106***	-3.097***
	(0.631)	(0.627)	(0.622)	(0.626)	(0.625)	(0.619)
Unmarried Partner	-5.414***	-5.434***	-5.541***	-5.443***	-5.387***	-5.579***
	(1.178)	(1.196)	(1.178)	(1.175)	(1.180)	(1.173)
Single Parent	1.908***	1.951***	1.903***	1.646***	1.835***	1.589***
	(0.576)	(0.561)	(0.561)	(0.567)	(0.569)	(0.570)
Interacted with Rank (CHKS Predicted in (2) and (5))						
Black	0.052**	0.083	0.049**	0.050**	0.124**	0.054**
	(0.021)	(0.055)	(0.020)	(0.022)	(0.052)	(0.021)
Hispanic	-0.010	0.048	0.002	-0.002	0.072	0.006
	(0.034)	(0.104)	(0.032)	(0.031)	(0.104)	(0.031)
Female	-0.045***	-0.112***	-0.054***	-0.046***	-0.115***	-0.055***
	(0.015)	(0.041)	(0.013)	(0.015)	(0.040)	(0.013)
Most Educated Parent*Rank						
< High School	0.037	0.126*	0.044*	0.043	0.161**	0.047*
	(0.027)	(0.074)	(0.027)	(0.026)	(0.071)	(0.026)
College+	-0.033**	-0.116***	-0.039**	-0.039**	-0.143***	-0.039**
	(0.016)	(0.044)	(0.015)	(0.016)	(0.043)	(0.016)
Forecast Causal Estimate			7.73**			13.21***
			(3.28)			(3.39)
Forecast Causal Estimate*Rank			0.180***			0.118*
			(0.053)			(0.062)
Constant	44.97***	20.00***	44.98***	45.51***	19.29***	46.01***
	(0.85)	(2.30)	(0.85)	(0.89)	(2.44)	(0.86)
R-Squared	0.11	0.11	0.12	0.11	0.12	0.12
Observations	49,725	49,248	49,232	49,725	49,248	49,232

In this table, I test the weighted OLS regressions of child rank on parent rank (models (1), (2) and (3)) and on predicted child rank (models (4), (5), and (6)). The predicted child rank is derived from the CHKS intercept (α_c) and slope (β_c) coefficients from the CZ-level child rank on parent rank regression as $y_{ic} = \alpha_c + \beta_c x_{ic} + e_{ic}$. The regressions test whether the spatial heterogeneity in mobility biases coefficients for individual and family characteristics. Each regression uses the cohort weights discussed in Appendix A with errors clustered at the CZ level. The CZ characteristics are the five primary ones found by CHKS to be most correlated with mobility: the spatial mismatch in access to jobs (fraction with < 15 minute commute), inequality (the Gini coefficient of the bottom 99%), school quality (measured by the high school dropout rate), social capital (index from Goetz and Rupasingha (2006)), and the fraction of single mothers. ***, **, * represent significance at the 0.01%, 0.05% and 0.10% level.

Table 6: Correlation Between Family and Demographic Characteristics and CHKS Mobility

	CZ Groups										1990 Census	
	5		10		20		25		50			
CHKS Coefficient	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Black	-0.93	0.90	-0.91	0.89	-0.81	0.87	-0.78	0.84	-0.75	0.81	-0.58	0.62
Hispanic	0.86	-0.91	0.70	-0.78	0.31	-0.49	0.28	-0.48	0.28	-0.47	0.07	-0.23
Parent Education												
< High School	0.17	-0.27	0.07	-0.20	-0.34	0.19	-0.39	0.21	-0.45	0.32	-0.52	0.43
High School	-0.88	0.93	-0.81	0.91	-0.72	0.88	-0.63	0.77	-0.61	0.69	-0.18	0.33
Some College	0.89	-0.90	0.63	-0.65	0.57	-0.68	0.63	-0.68	0.58	-0.62	0.52	-0.49
College +	0.74	-0.61	0.52	-0.41	0.73	-0.61	0.62	-0.53	0.58	-0.51	0.32	-0.40
Teen Parent	-0.97	0.94	-0.95	0.92	-0.92	0.90	-0.90	0.87	-0.81	0.81	-0.74	0.71
Single Parent	-0.82	0.81	-0.78	0.79	-0.85	0.76	-0.84	0.75	-0.80	0.74	-0.81	0.64
Unmarried Partner	-0.06	-0.07	-0.14	0.03	-0.28	0.15	-0.30	0.14	-0.23	0.13		

Low Mobility **High Mobility**

This table shows the correlation between the baseline model family and demographic characteristics and CHKS CZ-level mobility. In the first 10 columns, the CZs are grouped into k quantiles from greatest relative mobility to least and the characteristics are averaged over all CSD observations in each group. In the last two columns, the family characteristics are averaged within each cz for all children in the 1990 long form census. The correlations shown are unweighted correlations with the CZ intercept and slope for the 1990 census and with the observation-weighted CHKS intercept α_c^{CHKS} and slope β_c^{CHKS} from the CZ-level mobility regression in equation (3.3) for the CZ groups. The cells are highlighted from red to blue, where darker red represents a greater correlation with low mobility and darker blue represents a greater correlation with high mobility.

Table 7: Variation Explained by Family and Demographic Characteristics

Measure	CZ Quantile Groups 5	10	20	25	50	1990 Census Estimates
Regression Coefficient						
Intercept ($\alpha_{p,adjusted} = \gamma_\alpha + \delta_\alpha \alpha_p$)	0.44	0.47	0.54	0.55	0.65	0.61
Slope ($\beta_{p,adjusted} = \gamma_\beta + \delta_\beta \beta_p$)	0.58	0.54	0.61	0.67	0.72	0.72
Reduction in Dispersion						
Variance						
Intercept	0.74	0.72	0.66	0.62	0.47	0.50
Slope	0.63	0.66	0.52	0.42	0.37	0.36
Coefficient of Variation						
Intercept	0.54	0.53	0.48	0.45	0.35	0.41
Slope	0.31	0.34	0.21	0.14	0.09	-0.04
Mean Absolute Deviation						
Intercept	0.56	0.50	0.43	0.43	0.27	0.30
Slope	0.45	0.42	0.33	0.26	0.22	0.23

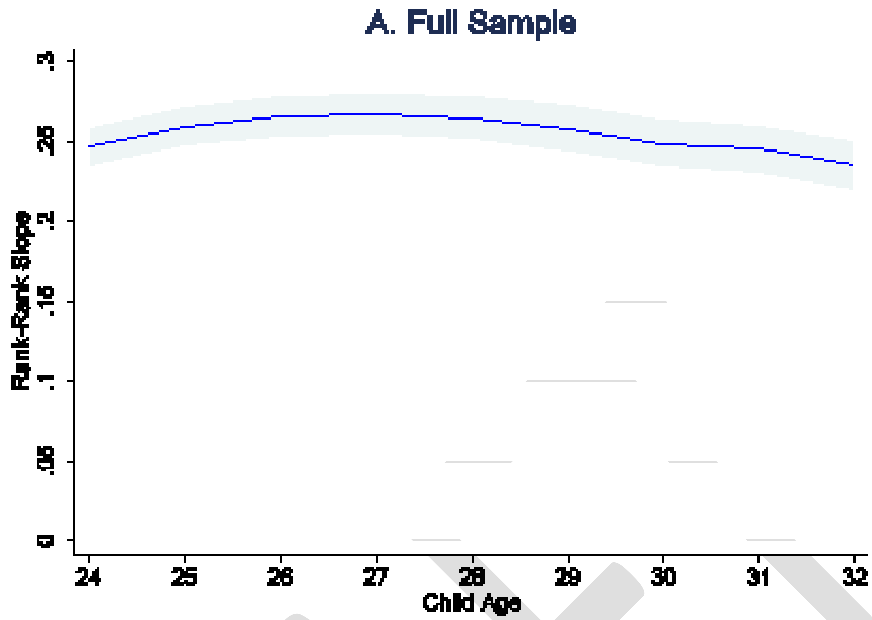
In this table, I test the variation explained by the demographic and family characteristics. In the first five columns, I divide the individuals into k ($p = 1, \ldots, k$) CZ quantile groups based on the relative mobility in each CZ. In the sixth column, I use the 1990 decennial census long form and calculate the predicted rank for each child using the CHKS CZ-level estimates. For columns 1-5, I create an adjusted rank for the CSD observations by taking the coefficients from the baseline model (model (1) in Table 5) and calculating $y_{i,Adjusted} = y_i - \beta_h h_i - \beta_{xh} x_i h_i$. I do the same to calculate an adjusted predicted rank for the 1990 census sample based on their family and demographic characteristics. The parent rank in the 1990 census was imputed by comparing their total family income in 1989 to all parents in the sample in the same older parent age cohort. The $y_{i,adjusted}$ accounts for the relationship between the observable characteristics and child rank but not parent rank. I then calculate the CZ-group level (for the CZ quantile groups) and CZ level (for the 1990 census sample) slope and intercept from the regression of adjusted child rank and predicted rank on parent rank. To measure how the demographic adjustment affects the CZ-group and CZ level estimates of mobility, I compare the adjusted to unadjusted coefficients in a variety of ways . First, I regress adjusted rank-rank slope on the unadjusted, and the adjusted rank-rank intercept on the unadjusted for each group or CZ. I also calculate a variety of dispersion measures for both intercept and slope coefficients: including variance, coefficient of variation, and mean absolute deviation. For each dispersion measure, I calculate the share of the variation explained by the observable characteristic as the reduction in dispersion in the adjusted measure compared to the unadjusted ($1 - \frac{\text{adjusted}}{\text{unadjusted}}$).

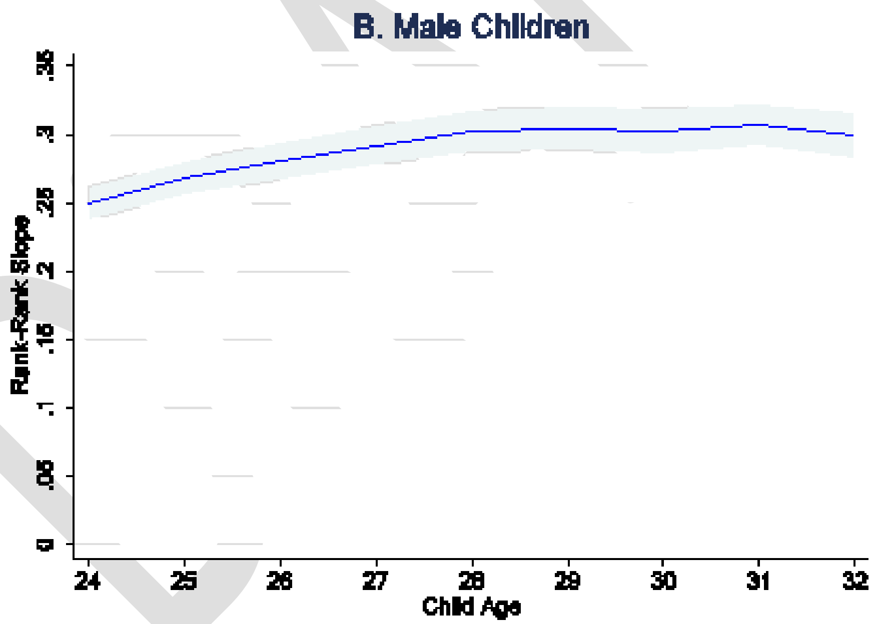
Appendix Table 1: Intergenerational Earnings Mobility and Family Characteristics

Dependent Variable = Child Rank	**Race and Parent Rank**	**Parent Rank Interactions**	**Baseline Model**
Variable	(1)	(2)	(3)
Parent Rank	0.219***	0.215***	0.209***
	(0.009)	(0.015)	(0.013)
Black	-10.15***	-15.17***	-15.64***
	(0.90)	(1.28)	(1.11)
Hispanic	-1.69	1.68	1.66
	(1.36)	(1.49)	(1.49)
Female		-8.24***	-8.39***
		(0.87)	(0.83)
Black*Female		14.09***	15.06***
		(1.54)	(1.08)
Most Educated Parent			
< High School		-6.55***	-6.74***
		(1.06)	(1.03)
Some College		2.34***	1.91***
		(0.89)	(0.47)
College+		8.38***	8.18***
		(1.06)	(1.05)
Teen Parent		-3.15***	-3.16***
		(0.63)	(0.63)
Unmarried Partner		-5.40***	-5.41***
		(1.18)	(1.18)
Single Parent		1.91***	1.91***
		(0.58)	(0.58)
Interacted with Parent Rank			
Black	0.075***	0.038	0.052**
	(0.021)	(0.035)	(0.021)
Hispanic	0.007	-0.011	-0.010
	(0.032)	(0.034)	(0.034)
Female		-0.048***	-0.045***
		(0.016)	(0.015)
Black*Female		0.028	
		(0.045)	
Most Educated Parent*Rank			
< High School		0.033	0.037
		(0.028)	(0.027)
Some College		-0.009	
		(0.018)	
College+		-0.038**	-0.033**
		(0.017)	(0.016)
Constant	41.45***	44.70***	44.97***
	(0.58)	(0.89)	(0.85)
R-Squared	0.07	0.11	0.11
Observations	49,725	49,725	49,725

In this table, I regress child rank on parent rank and a variety of other demographic and family characteristics using the cohort weights discussed in Appendix A. In, model (1), I include race dummies and race interacted with parent rank. In model (2), I add a richer set of family characteristics as dummies and interacted with parent rank. In model (3), the baseline model, I include the less than high school and college+ education interactions as both are significant in the either in the weighted or unweighted regressions (and nearly so in the other) and the point estimates for both are large in magnitude. The errors are clustered at the CZ level. ***, **, * represent significance at the 0.01%, 0.05% and 0.10% level.

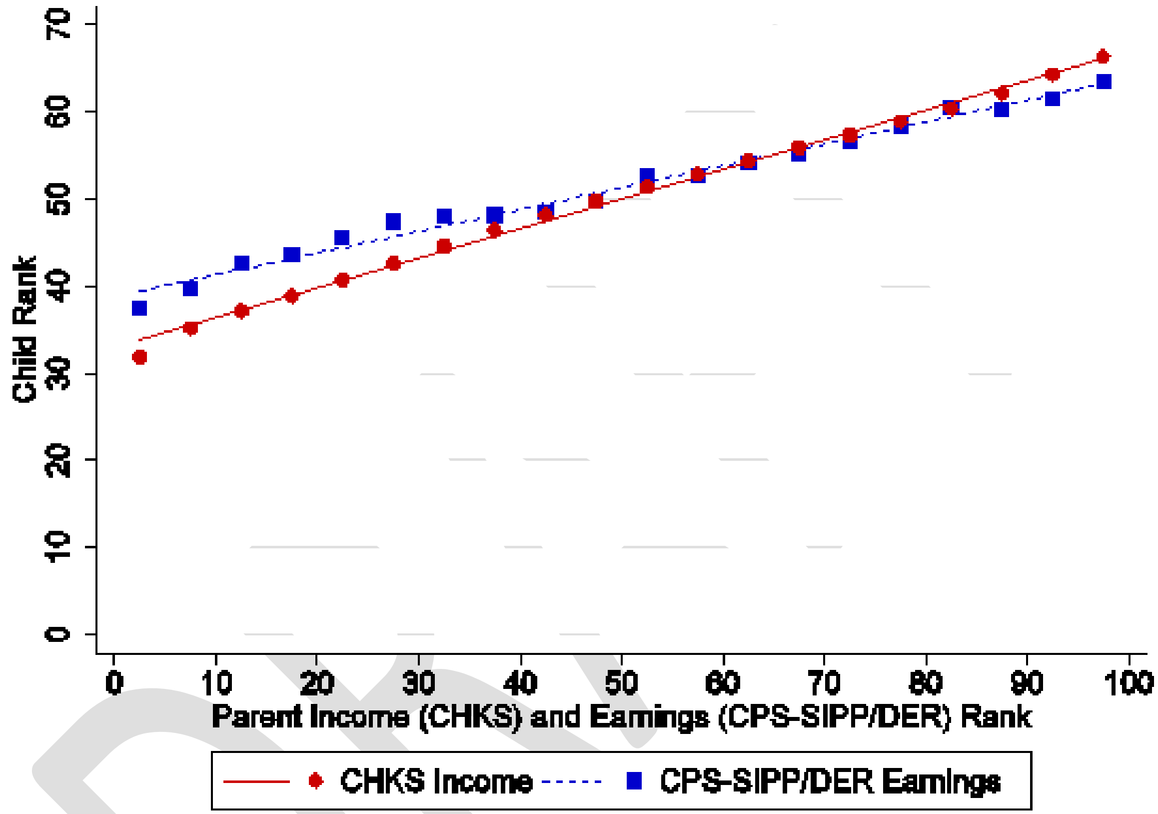
Figure 1: Intergenerational Mobility by Age of Child Earnings Measurement





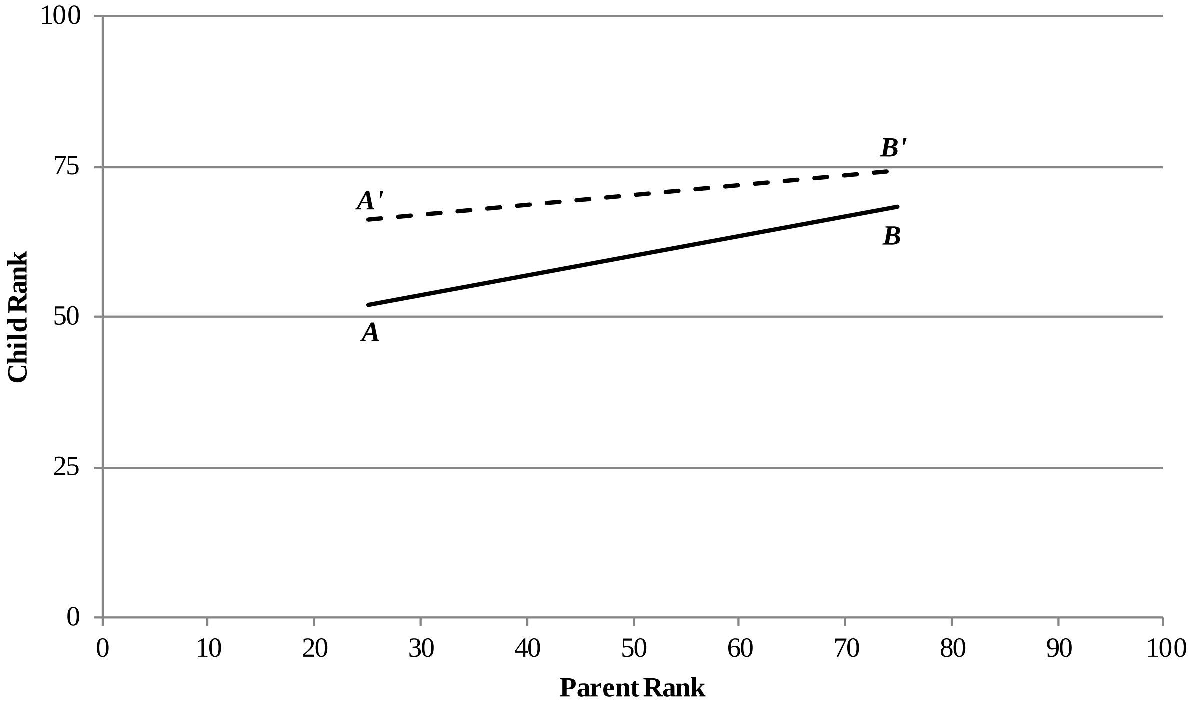
These figures plot the rank-rank slope of intergenerational mobility. The parent family earnings are the average when the older parent is 40-44 years old. The child earnings are the average of earnings for age t and $t + 1$ where t varies from 24 to 32. It should be noted that as t varies, so does the size of the sample because more children in the CPS ASEC and SIPP from 1991 on reach $t + 1$ by 2011, the last available year of DER earnings data. On average for each year younger of t, the sample increases by about 18%. For example, at $t = 32$, there are 28,798 parent-child pairs and at $t = 24$, there are 106,594 parent-child pairs. Panel A shows the rank-rank slope by age for the full sample, and Panel B shows the slope by age for male children only

Figure 2: Association between Parent and Child Rank in CHKS and CPS-SIPP/DER



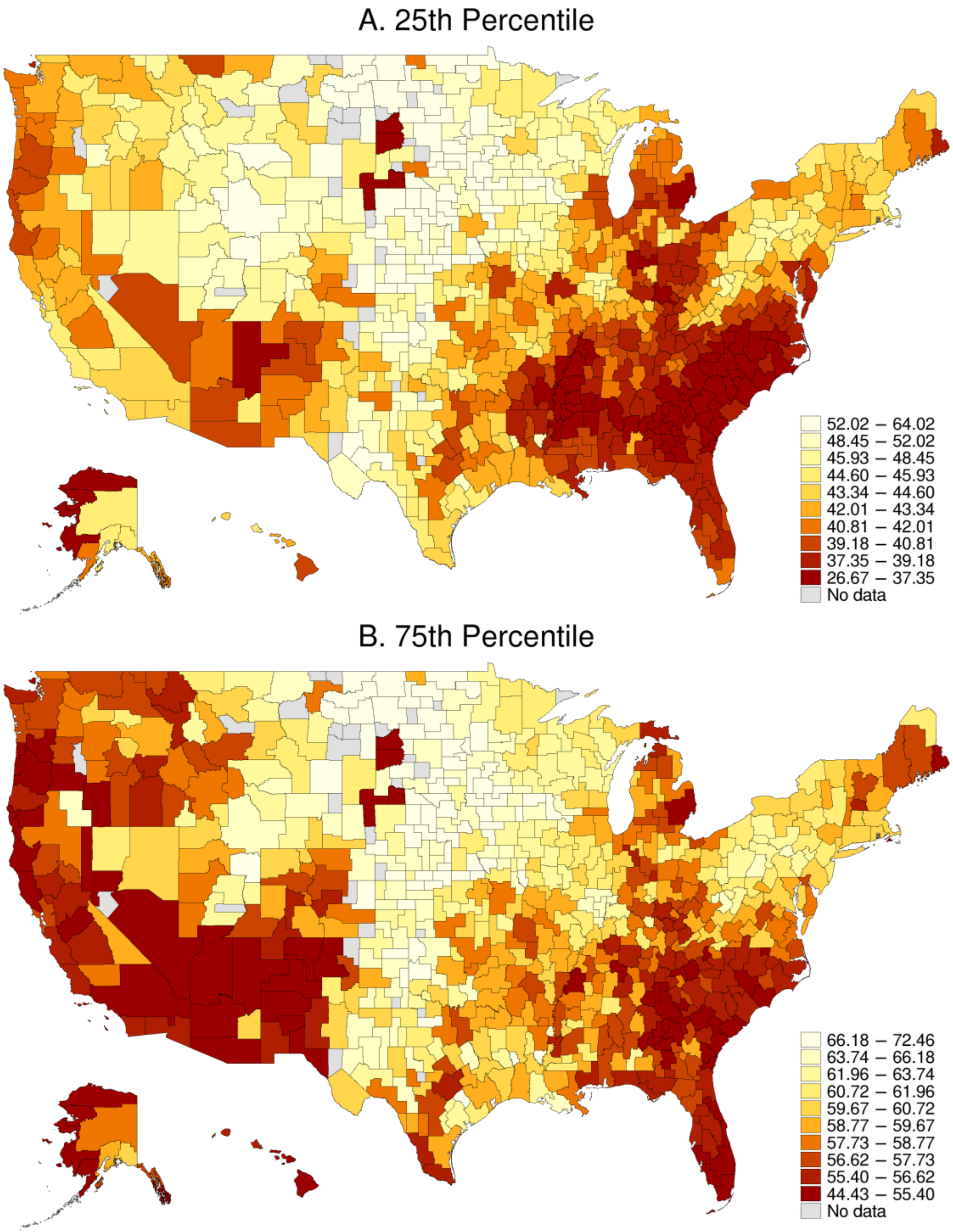
This figure plots the average child rank for 20 parent rank bins. In both the CPS-SIPP/DER and the CHKS data, the relationship between parent rank and average child rank is very well represented by the linear regression slope and intercepts on the individual observations. The CHKS parent ranks are determined by parent income rank from 1996-2000 and child ranks in 2011-2012 (child ages 29-32, depending on the birth year of 1980-1982). The CPS-SIPP/DER ranks are determined by ranking the earnings of each parent family against all parent families in the same age cohort (of the older parent) at ages 40-44. The CPS-SIPP/DER child ranks are determined by ranking each individual child against all individuals in the same age cohort, with earnings measured at 29-30 years old. The CHKS bins are calculated from data made available on their website at http://www.equality-of-opportunity.org/.

Figure 3: Example of the Parent Characteristics Adjustment



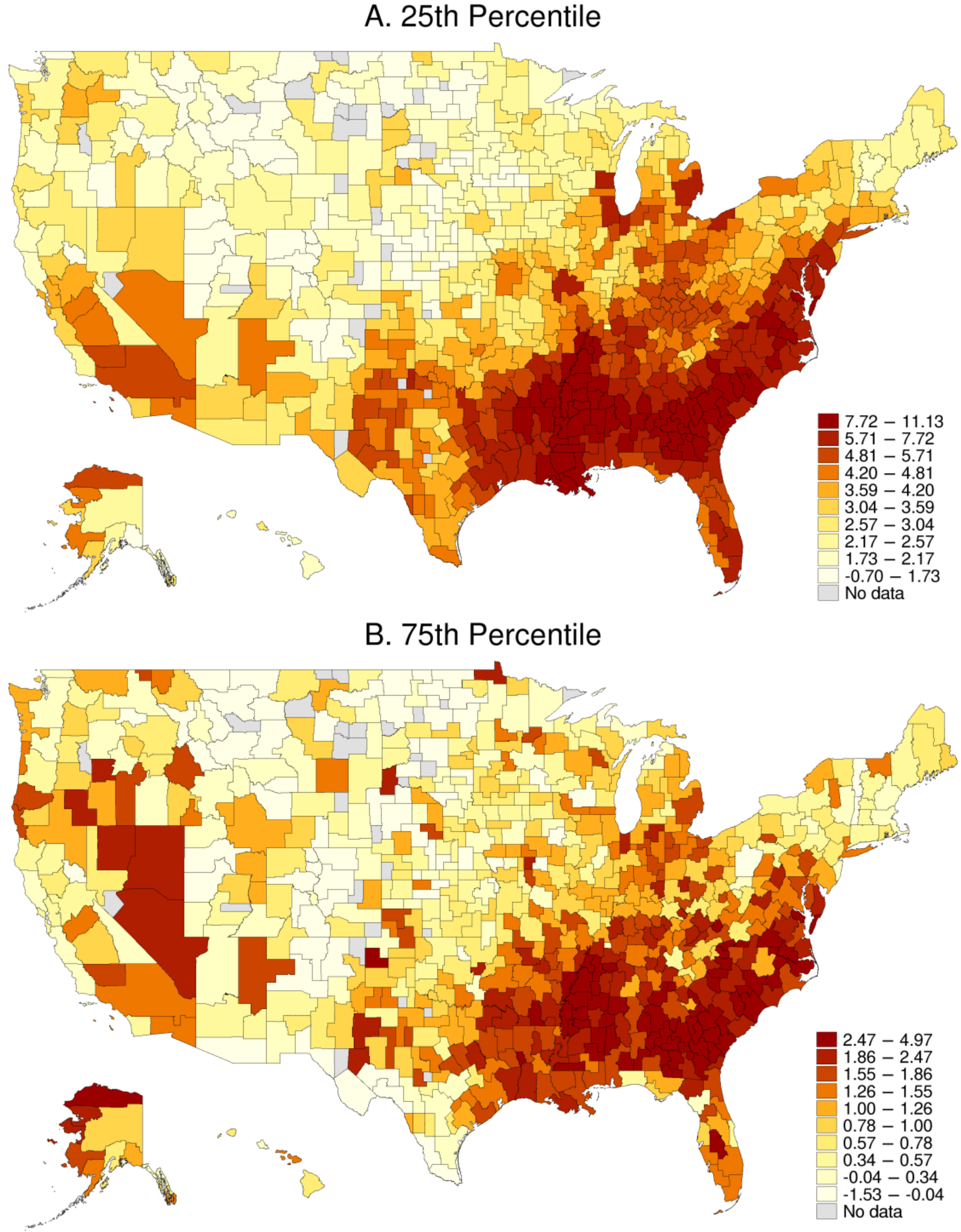
The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and using microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort and the child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed in the census, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$ and conducted the rank-rank regression of the adjusted child rank on the estimated parent rank, adjusting each CZ to the model baseline group of white sons of married, high school educated parents. In the New York CZ, the CHKS intercept and slope are 43.67 and 33.00 respectively. Suppose there are only two children in New York, A) a Black son of high school graduates at the 25th percentile and B) a White daughter of college graduates at the 75th percentile. For each, I assign a predicted income rank based on the CHKS coefficients, for A: $43.67 + 33(0.25) = 51.92$ and for B: $43.67 + 33(0.75) = 68.42$. I adjust the predicted rank for child A based on his characteristics as follows: 1) Black: $+\ 15.64$ and 2) Black*parent rank: $-0.052(25)$. The total adjustment for A is $15.64 - 0.052(25) = 14.34$, to get an adjusted rank A' of 66.26. For B, the adjustments are 1) female: $+8.39$, 2) female*parent rank: $+0.045(75)$, 3) college graduates: -8.18, and 4) college graduates*parent rank: $+0.033(75)$. The total adjustment for B is $8.39 + 0.045(75) - 8.18 + 0.033(75) = 6.06$ to get an adjusted rank B' of 74.48. The adjusted slope for New York would be $\frac{74.48-66.26}{0.75-0.25} = 16.44$ and the adjusted intercept would be $66.26 - 16.44(0.25) = 62.15$.

Figure 4: CHKS Expected Child Rank at 25th and 75th Percentile



This figure shows the geographic variation in the expected child rank at different parent income ranks across CZs. Panel A shows expected rank for a child with parents at the 25th percentile and Panel B shows the expected rank for a child with parents at the 75th percentile of the income distribution.

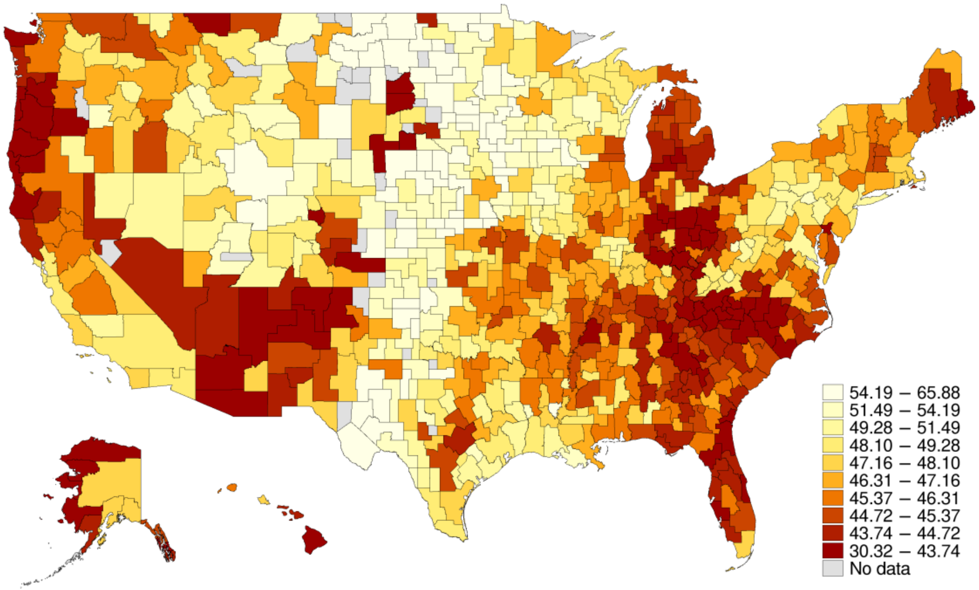
Figure 5: Impact of Parent Characteristic Adjustment on Child Rank at 25th and 75th Percentile



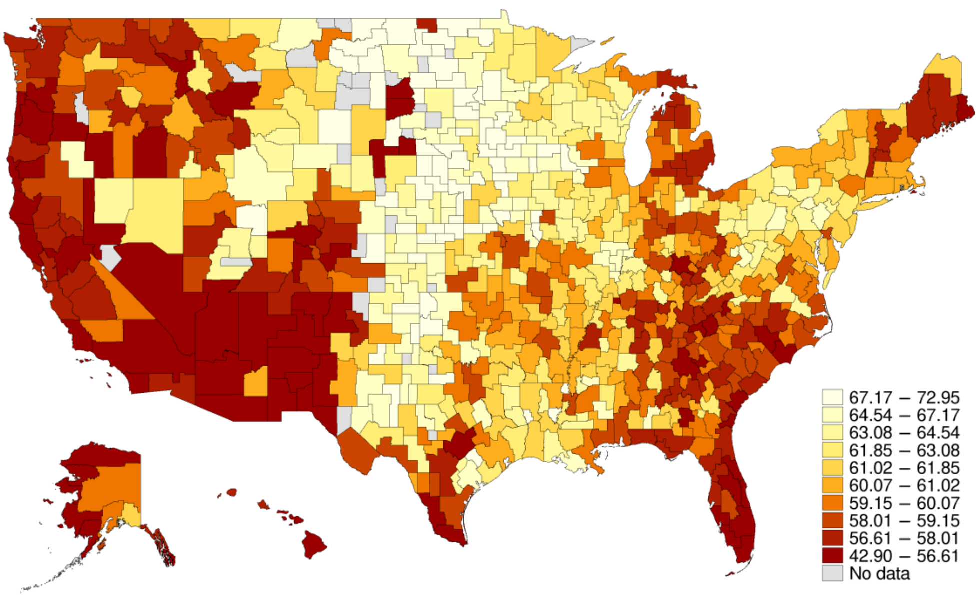
This figure shows how the expected child rank is affected by their family and demographic characteristics. Panel A shows the adjustment to the expected rank for a child with parents at the 25th percentile and Panel B shows the adjustment for a child with parents at the 75th percentile of the income distribution The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and using microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort and the child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed in the census, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$ and conducted the rank-rank regression of the adjusted child rank on the estimated parent rank, adjusting each CZ to the model baseline group of white sons of married, high school educated parents. In both panels, the adjustment is largest (dark red) in the Southeast and parts of the East Coast. A larger adjustment means that the family and demographic characteristics of children living there are associated with lower mobility. Both maps are divided into ten equally sized quantiles.

Figure 6: Expected Child Rank after Parent Characteristic Adjustment

A. 25th Percentile

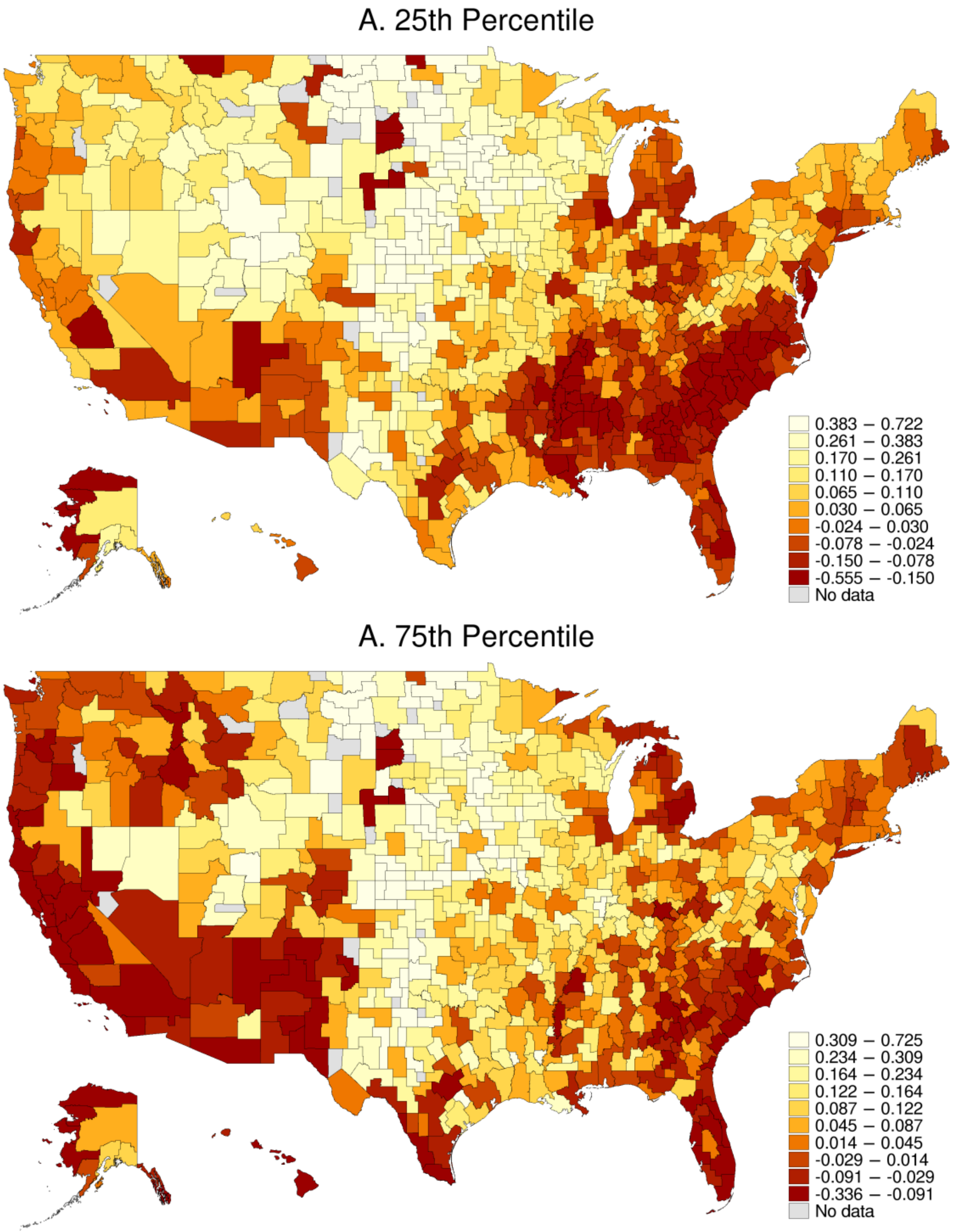


B. 75th Percentile



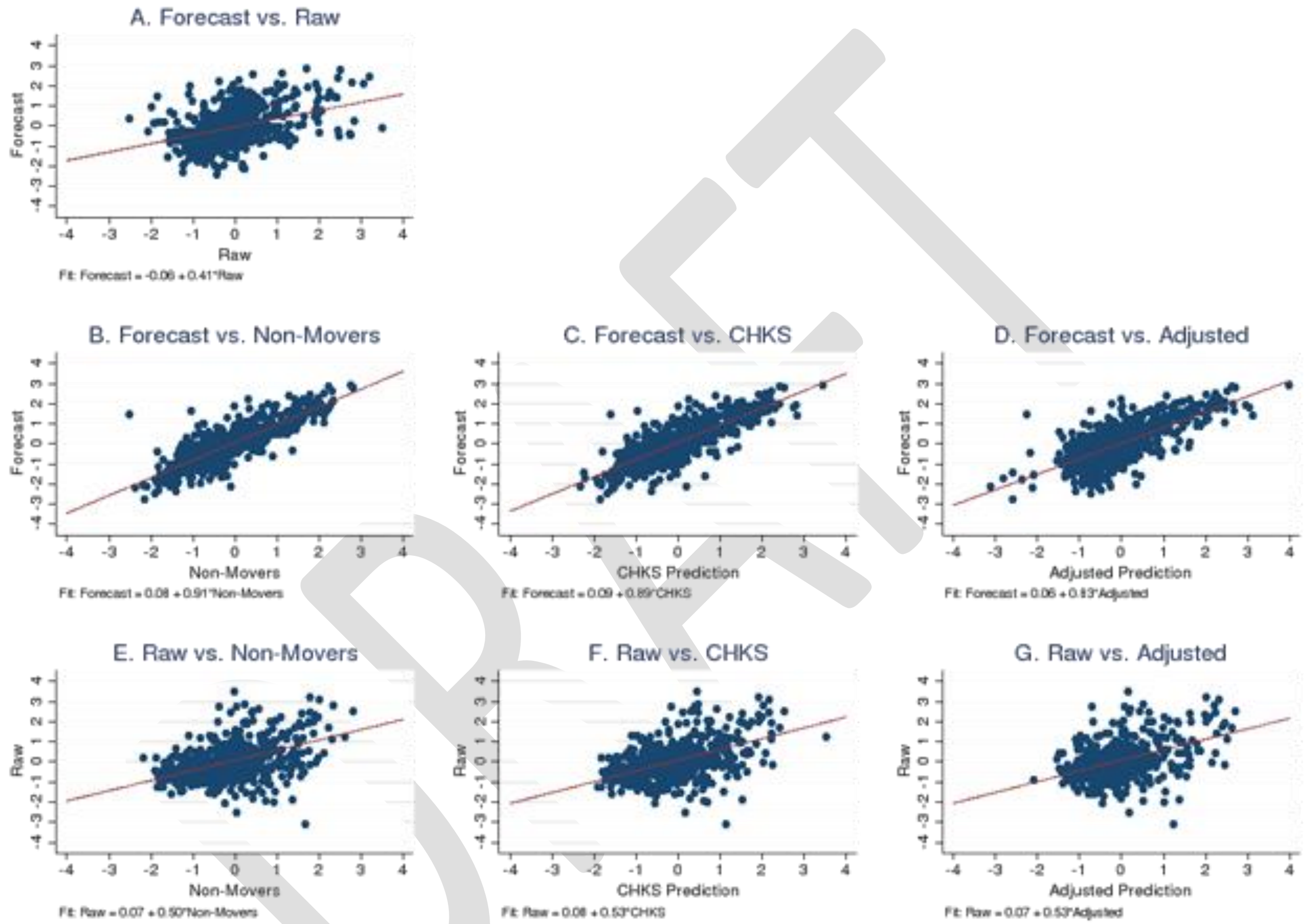
This figure shows how the expected child rank for children from below median (Panel A) and above median (Panel B) families after adjusting for the parent characteristics of the children who live there. The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort. Each child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$, adjusting each CZ to the model baseline group of white sons of married, high school educated parents. The CZ-level rank-rank regression was conducted with the adjusted child rank regressed on the estimated parent rank. Both maps are divided into ten equally sized quantiles. Compared to Figure 4, the South is no longer such an outlier. The unweighted correlation between the adjusted and unadjusted expected rank at the 25th and 75th percentiles are 0.93 and 0.73 respectively. Weighted by the 2000 population, the correlations are 0.86 and 0.47.

Figure 7: Causal Mobility Estimates from Chetty and Hendren



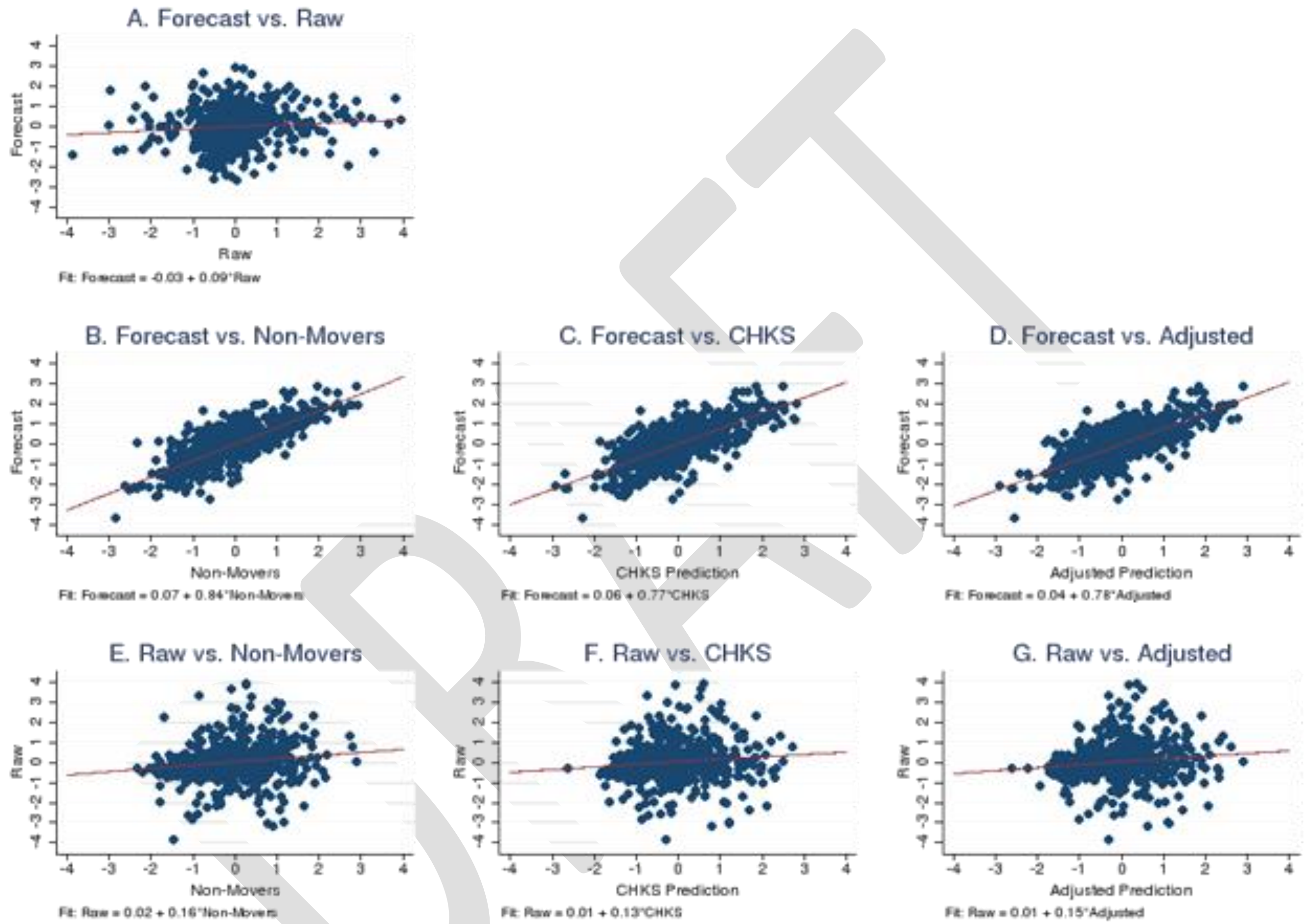
This figure shows how the forecast causal estimates from Chetty and Hendren for children from below median (Panel A) and above median (Panel B) families. The forecasts were created by taking the raw causal estimates and combining them with data on mobility of non-movers to address the fact that 71% of the variation in the raw causal estimates was due to sampling variation and not the causal effects of place. The weight given to non-movers in the forecast for each CZ is based on the precision of the raw causal estimate. Both maps are divided into ten equally sized quantiles.

Figure 8: Comparing 25th Percentile Causal Estimates with Predictions



This figure shows how the correlation between the standardized causal estimates and predictions by CZ for children from below median income families. In Panel A, the Chetty and Hendren raw causal estimates are compared to their forecast estimates that use results from non-movers to increase the precision of the raw estimates based on movers only. In Panels B, C, and D, the forecast estimates are compared to permanent residents, CHKS predictions, and parent and demographic characteristic adjusted predictions (adjusted from CHKS). Panels D, E, and F compare the raw causal estimates to the same three estimates as A-C. The forecast is highly correlated with each estimate. However, the raw causal estimates are not.

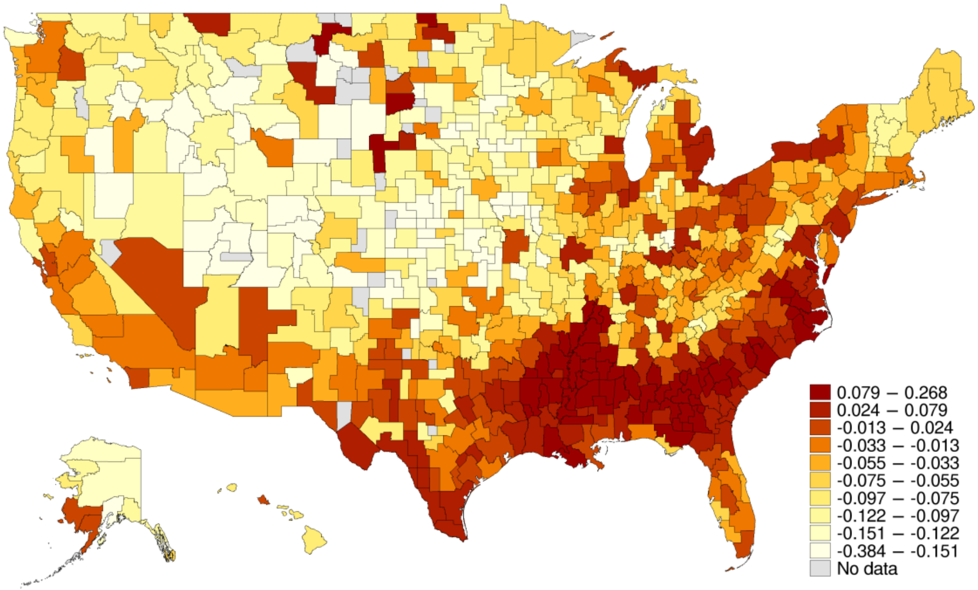
Figure 9: Comparing 75th Percentile Causal Estimates with Predictions



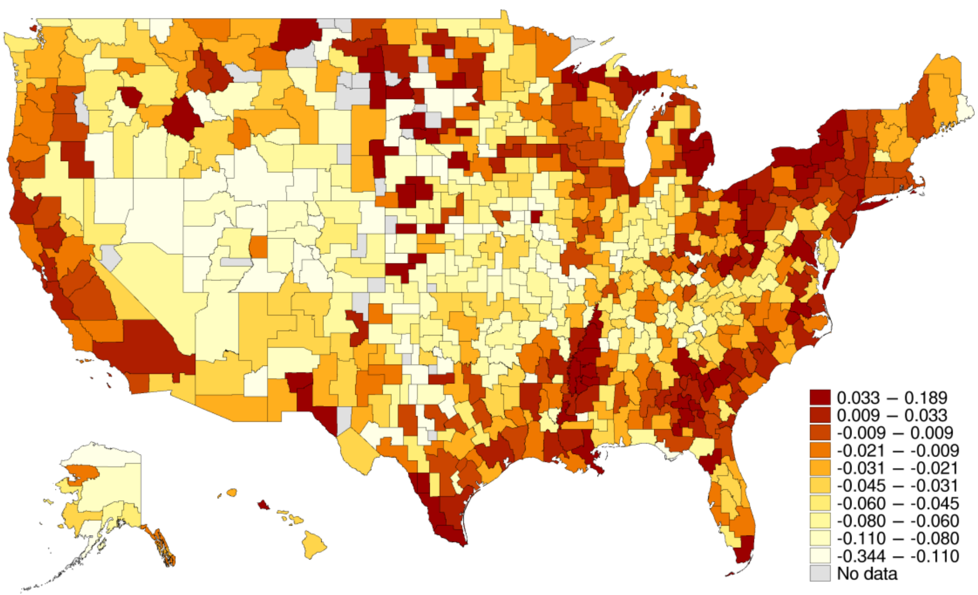
This figure shows how the correlation between the standardized causal estimates and predictions by CZ for children from above median income families. In Panel A, the Chetty and Hendren raw causal estimates are compared to their forecast estimates that use results from non-movers to increase the precision of the raw estimates based on movers only. In Panels B, C, and D, the forecast estimates are compared to permanent residents, CHKS predictions, and parent and demographic characteristic adjusted predictions (adjusted from CHKS). Panels D, E, and F compare the raw causal estimates to the same three estimates as A-C. The forecast is highly correlated with the permanent resident and adjusted estimates. However, the raw causal estimates have almost no correlation with any.

Figure 10: Impact of Family Characteristic Adjustment on Causal Mobility Estimates

A. 25th Percentile

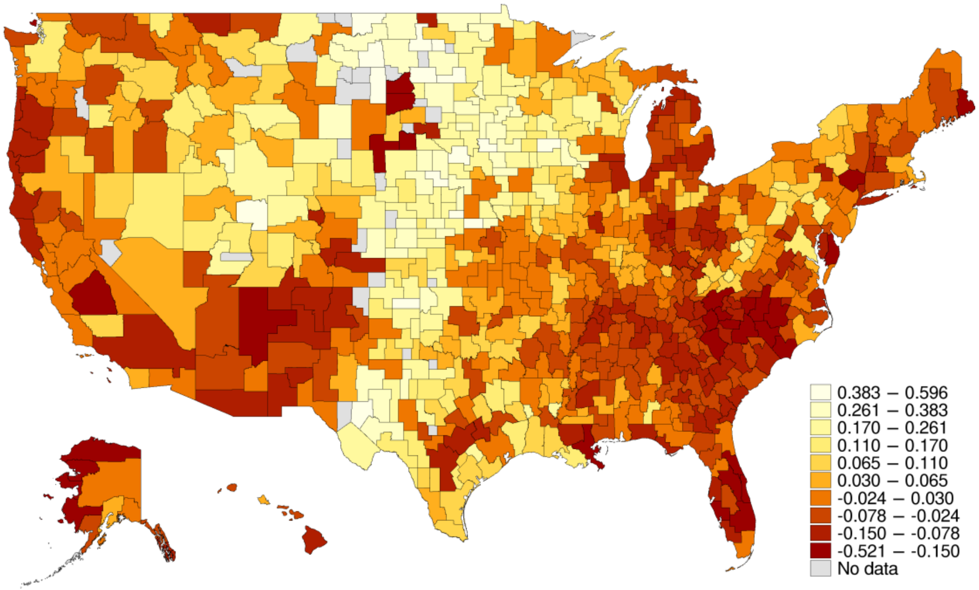


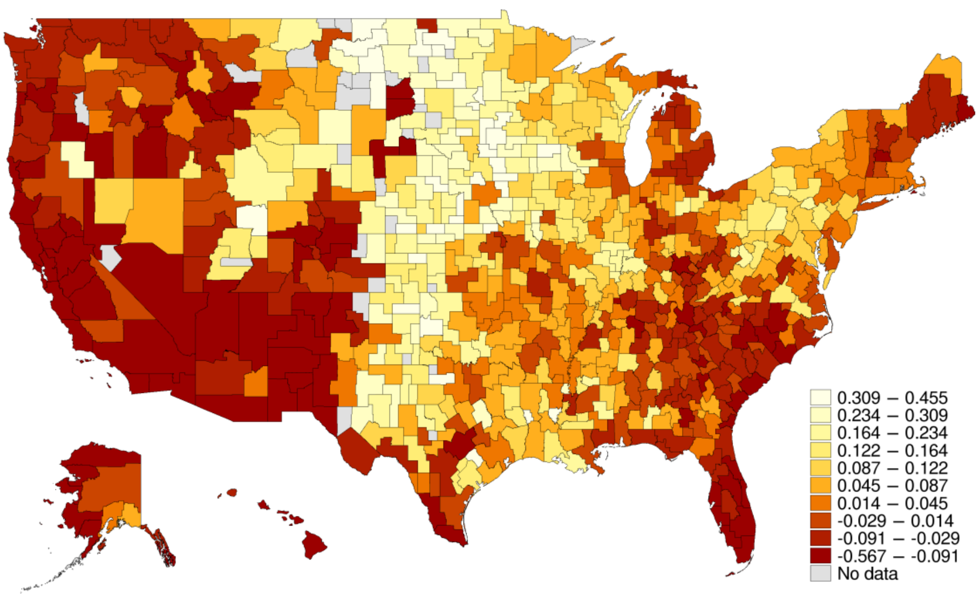
B. 75th Percentile



This figure shows how the forecast causal estimates from Chetty and Hendren is affected by replacing their estimates for non-mover mobility with the demographic and family characteristic adjusted estimates. The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and using microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort and the child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed in the census, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$ and conducted the rank-rank regression of the adjusted child rank on the estimated parent rank, adjusting each CZ to the model baseline group of white sons of married, high school educated parents. Both maps are divided into ten equally sized quantiles.

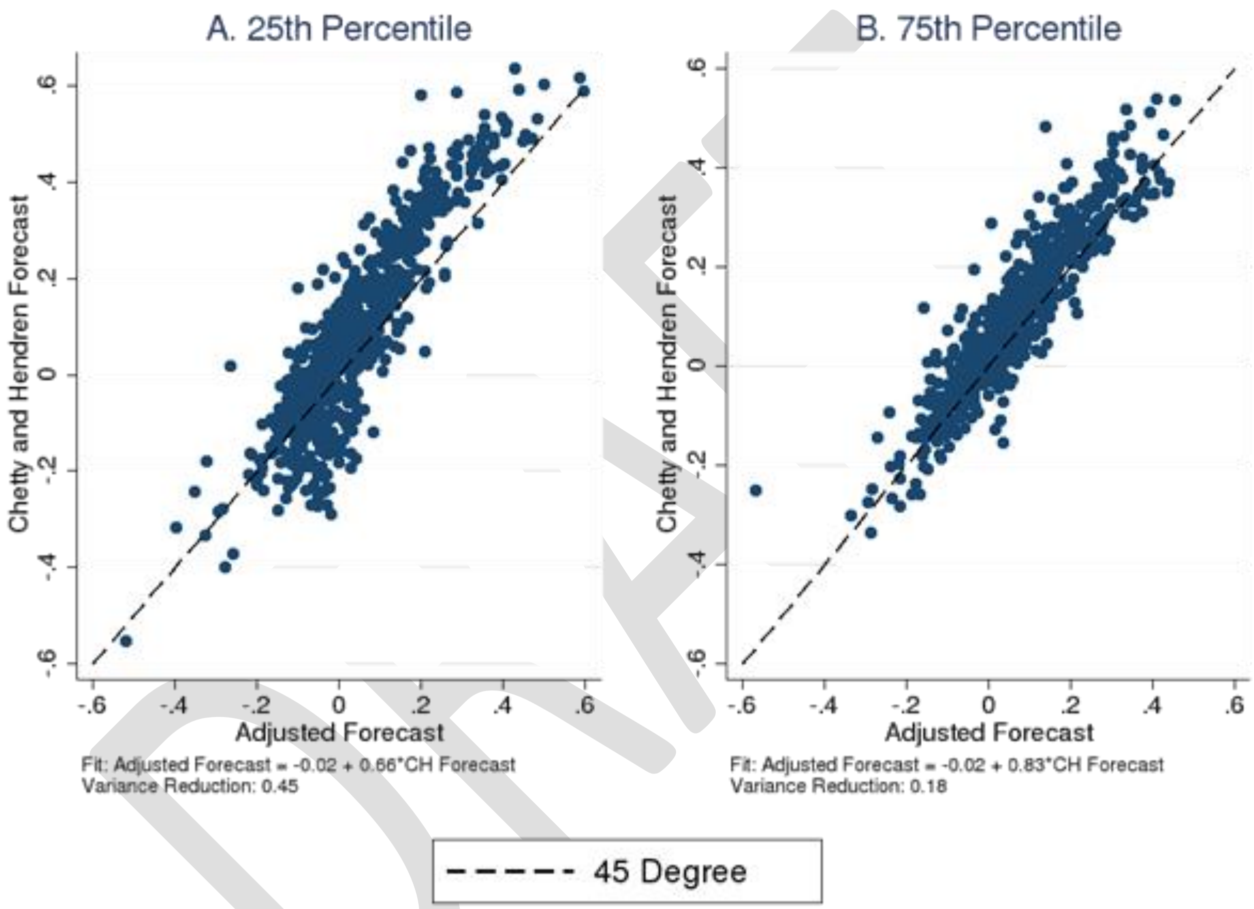
Figure 11: Causal Mobility Estimates Adjusted for Family Characteristics





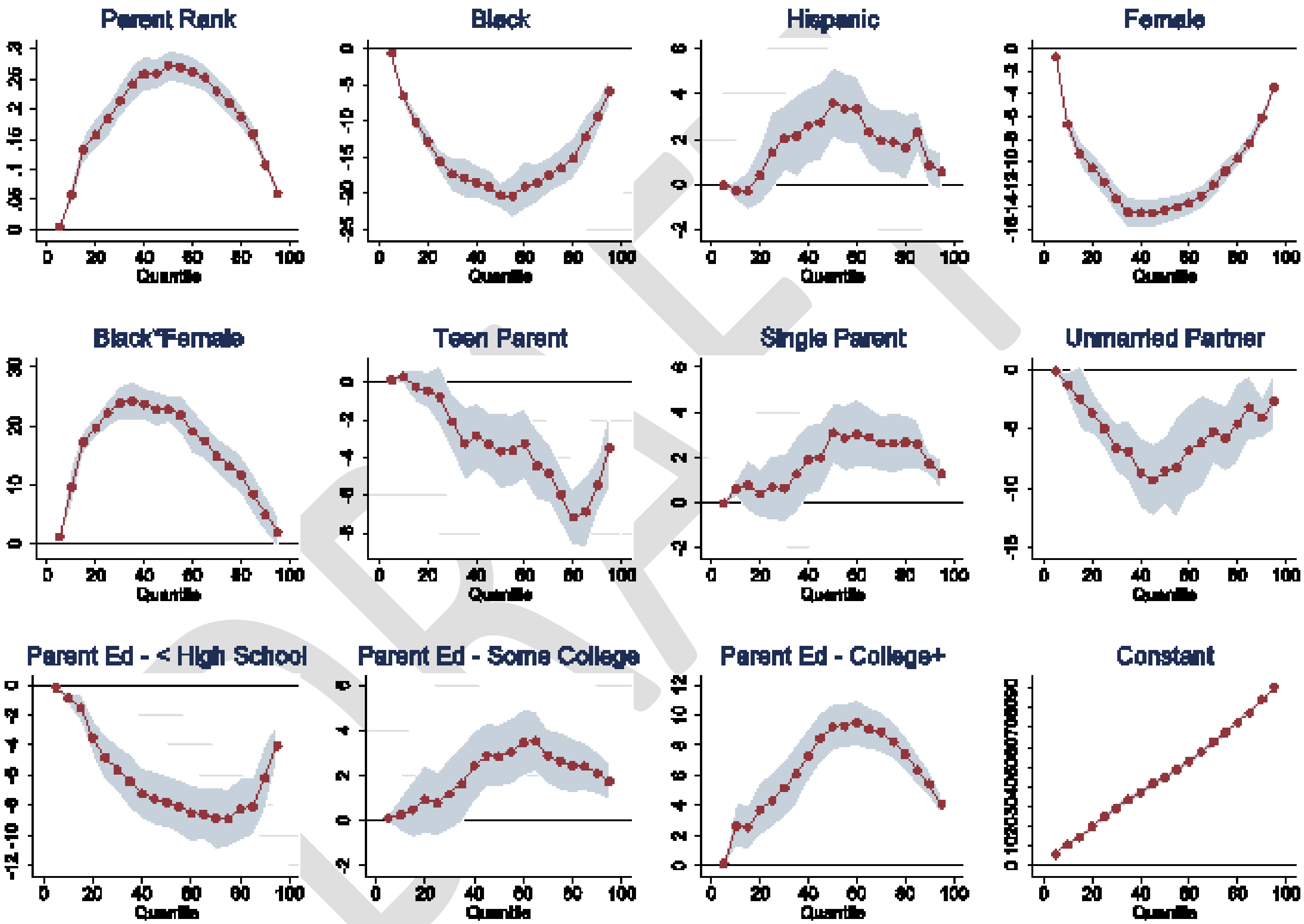
This figure shows how the forecast causal estimates from Chetty and Hendren is affected by replacing their forecast term for non-mover mobility with the demographic and family characteristic adjusted estimates (on the same scale as Figure 7). The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and using microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort and the child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed in the census, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$ and conducted the rank-rank regression of the adjusted child rank on the estimated parent rank, adjusting each CZ to the model baseline group of white sons of married, high school educated parents. Both maps are divided into ten equally sized quantiles.

Figure 12: Comparing Chetty and Hendren Causal Mobility Estimates to Family Characteristic Adjusted Estimates



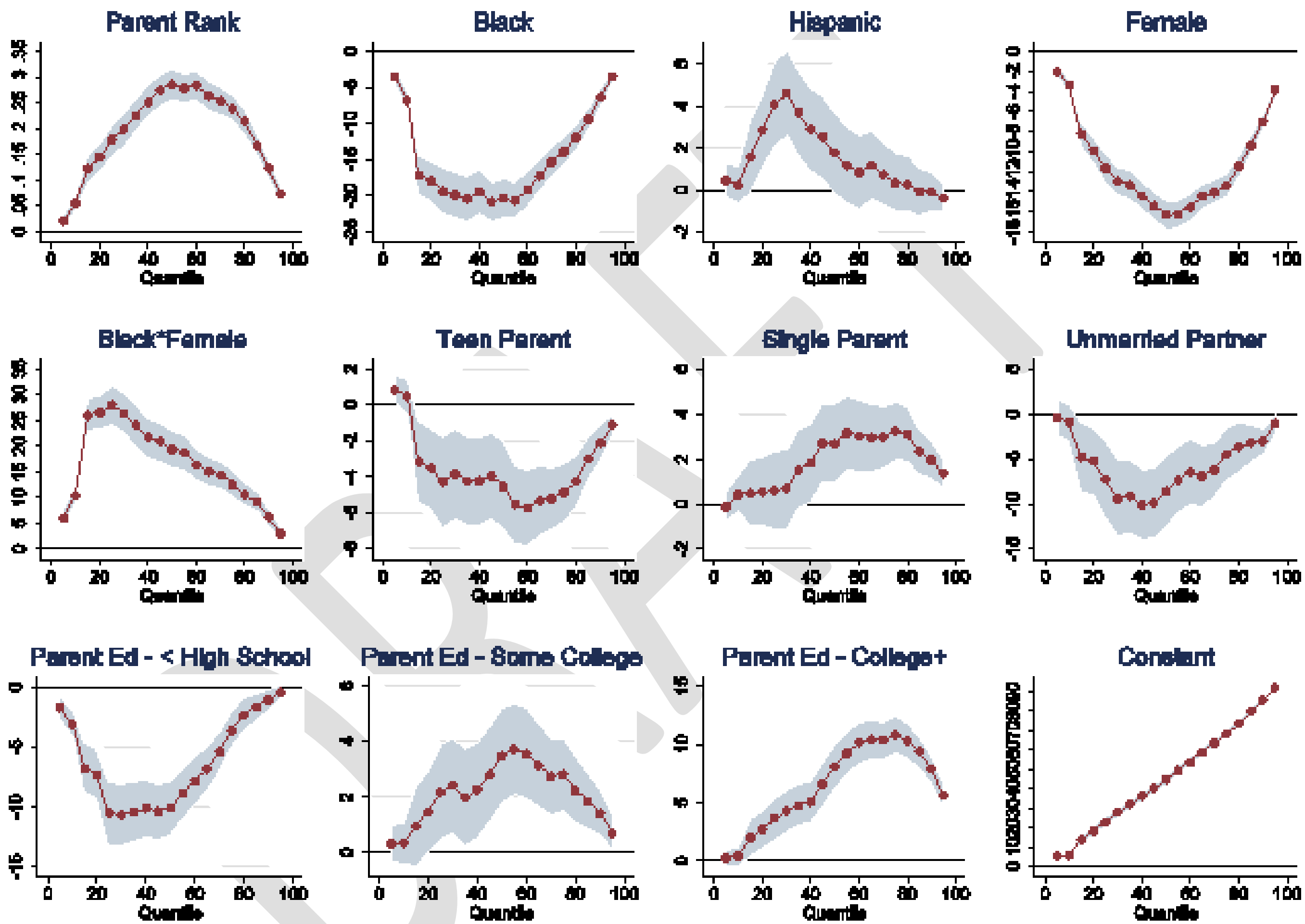
This figure compares the forecast causal estimates from Chetty and Hendren with and without the adjustment of the term for non -mover mobility for child demographic and family characteristic. The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and using microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort and the child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed in the census, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$ and conducted the rank-rank regression of the adjusted child rank on the estimated parent rank, adjusting each CZ to the model baseline group of white sons of married, high school educated parents.

Figure 13: Conditional Rank-Rank Quantile Regression on Family Characteristics



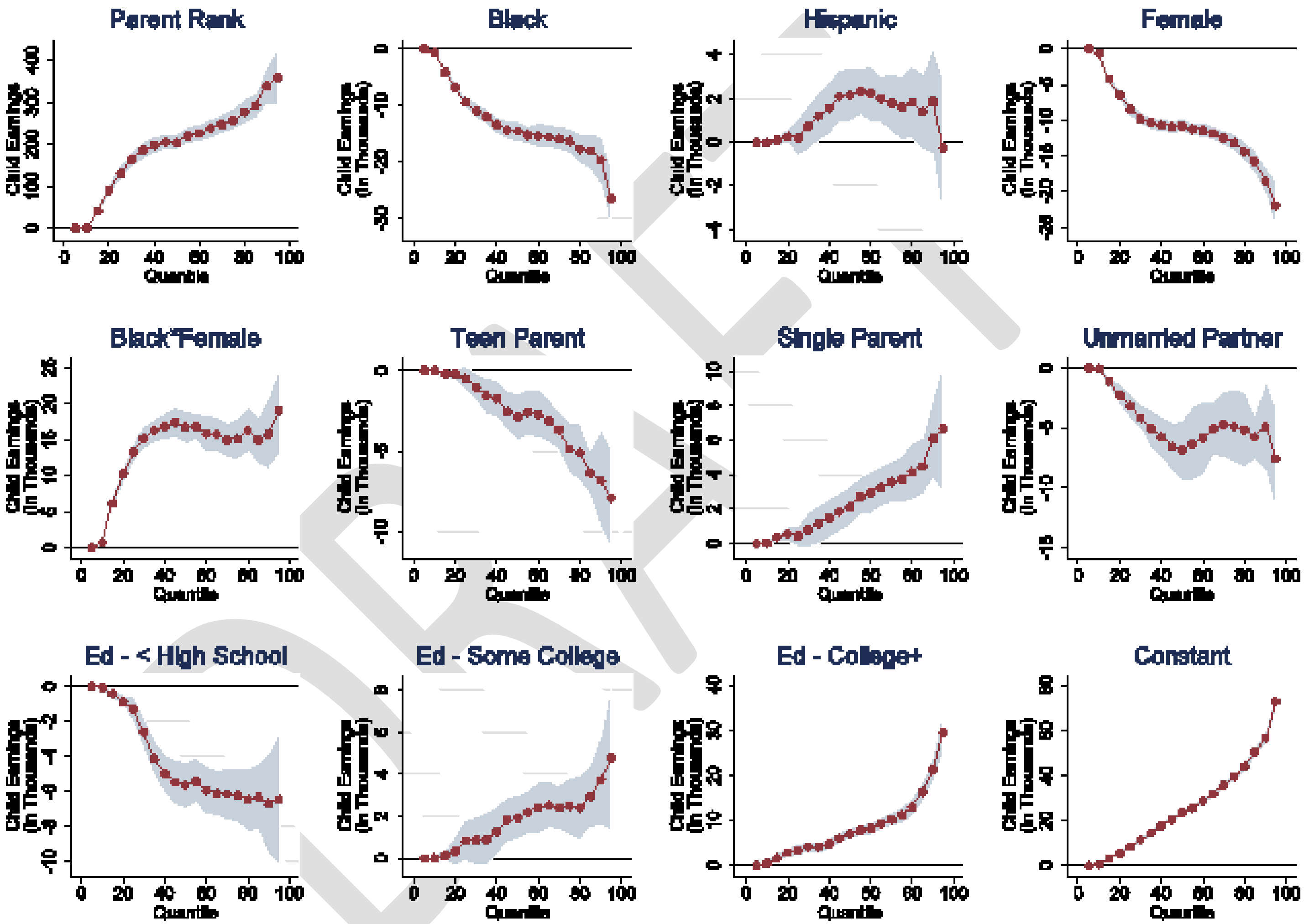
This figure plots the results of quantile regressions of child rank on parent rank and each of the family and demographic characteristics using cohort weights. The 95% confidence intervals are shown.

Figure 14: Unconditional Rank-Rank Quantile Regression on Family Characteristics



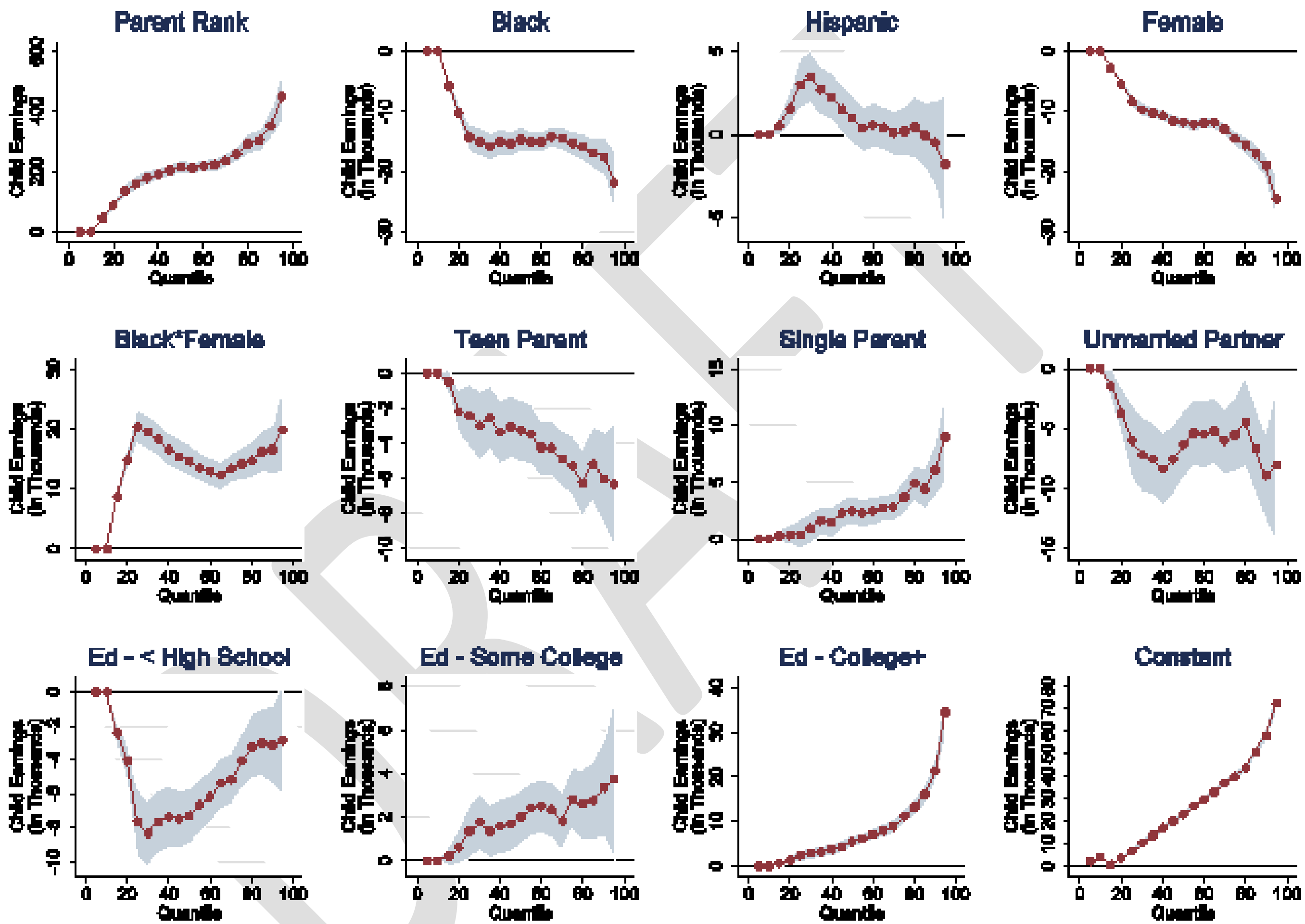
This figure plots the results of unconditional quantile regressions of child rank on parent rank and each of the family and demographic characteristics using cohort weights. This shows how the sample child distribution would differ if each characteristic where changed from the baseline (white, high school educated, married non-teen parents) to the characteristic in each subplot. This shows where in the distribution the lesser or greater of up ward mobility of different subgroups has an effect. The 95% confidence intervals are shown.

Figure 15: Conditional Earnings-Rank Quantile Regression on Family Characteristics



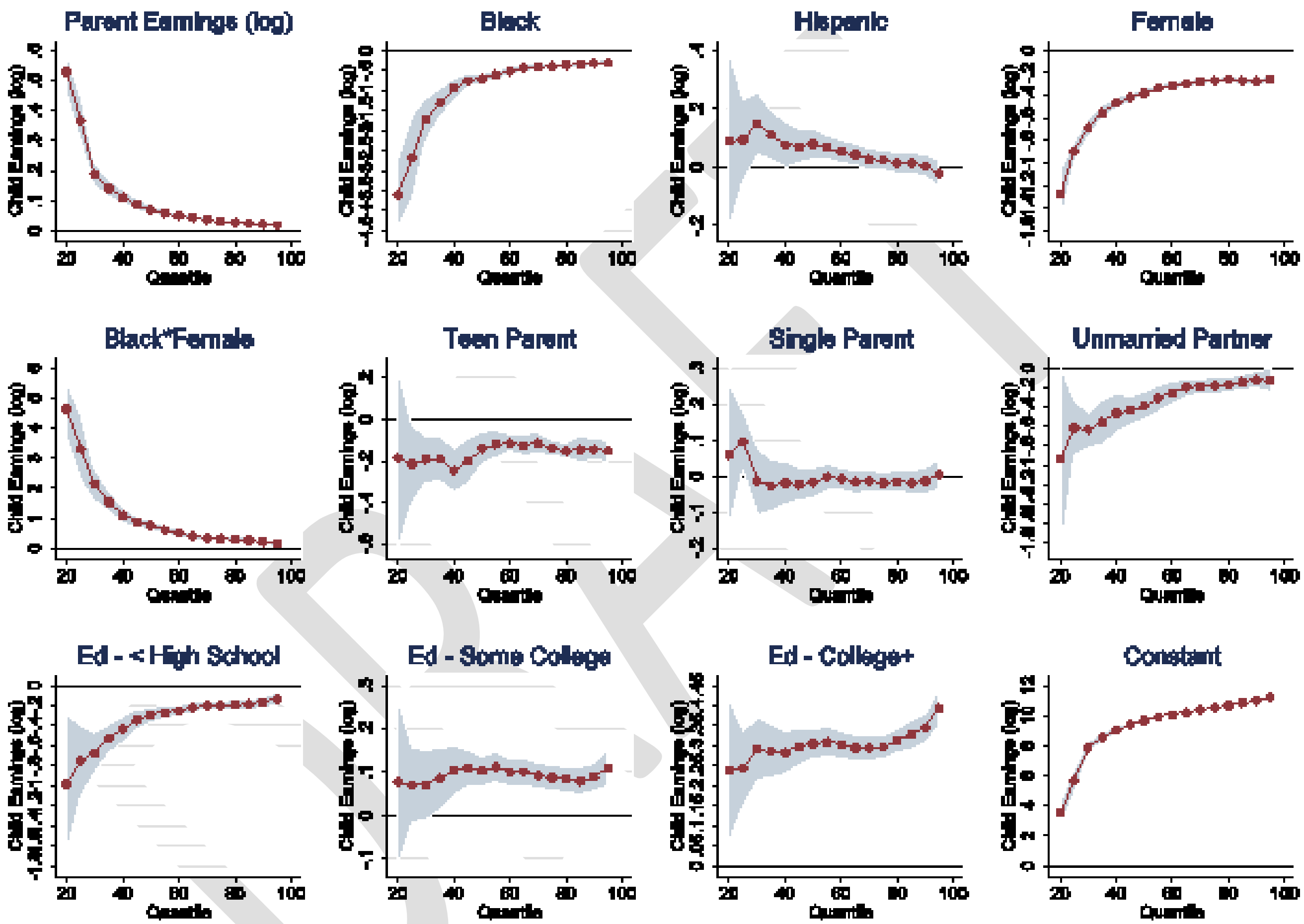
This figure plots the results of quantile regressions of child rank on parent rank and each of the family and demographic characteristics using cohort weights. The 95% confidence intervals are shown.

Figure 16: Unconditional Earnings-Rank Quantile Regression on Family Characteristics



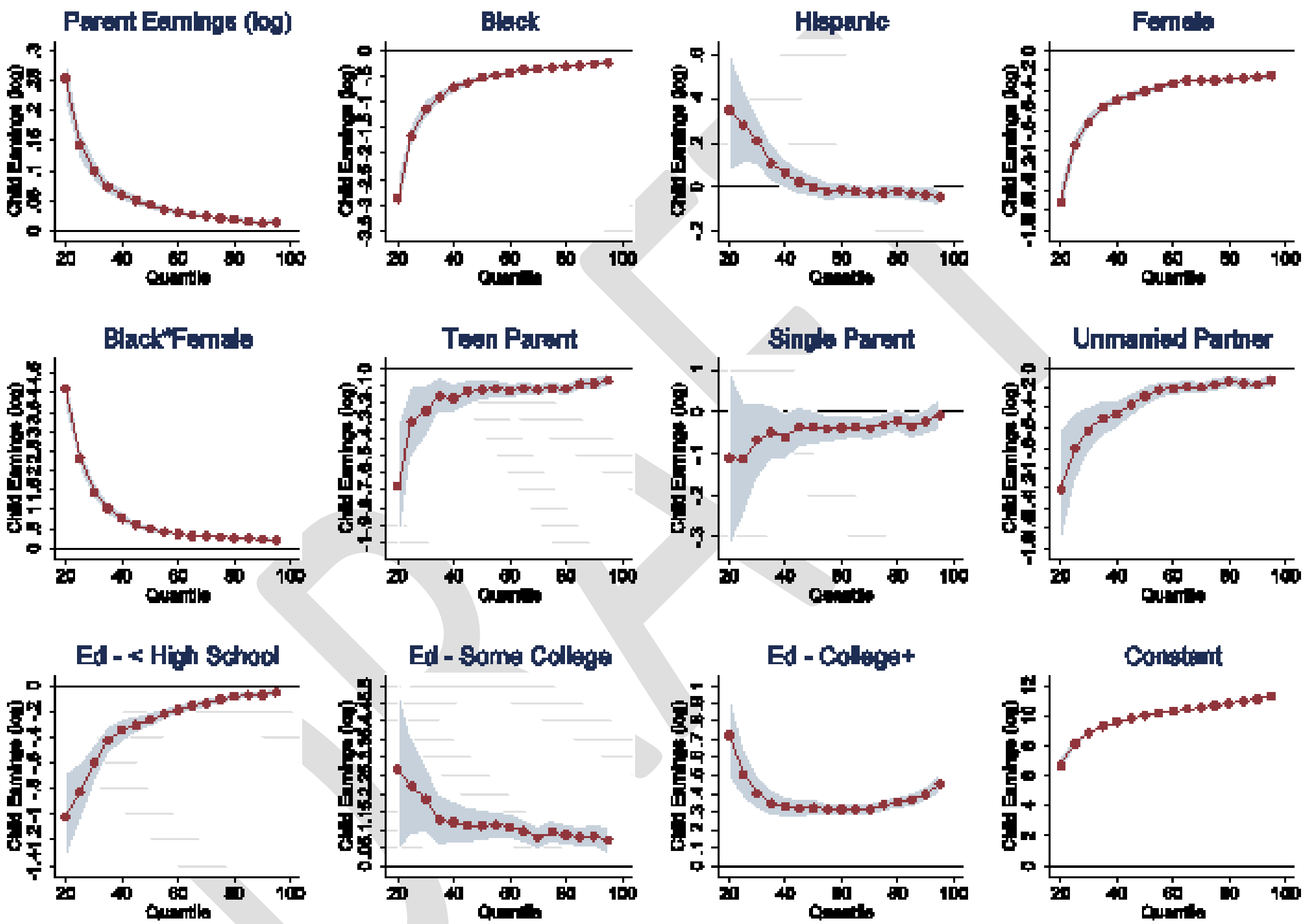
This figure plots the results of unconditional quantile regressions of child rank on parent rank and each of the family and d emographic characteristics using cohort weights. This shows how the sample child distribution would differ if each characteris tic where changed from the baseline (white, high school educated, married non-teen parents) to the characteristic in each subplot. This shows where in the distribution the lesser or greater of upward mobility of different subgroups has an effect. The 95% confidence intervals are shown.

Figure 17: Conditional Log Earnings Quantile Regression on Family Characteristics



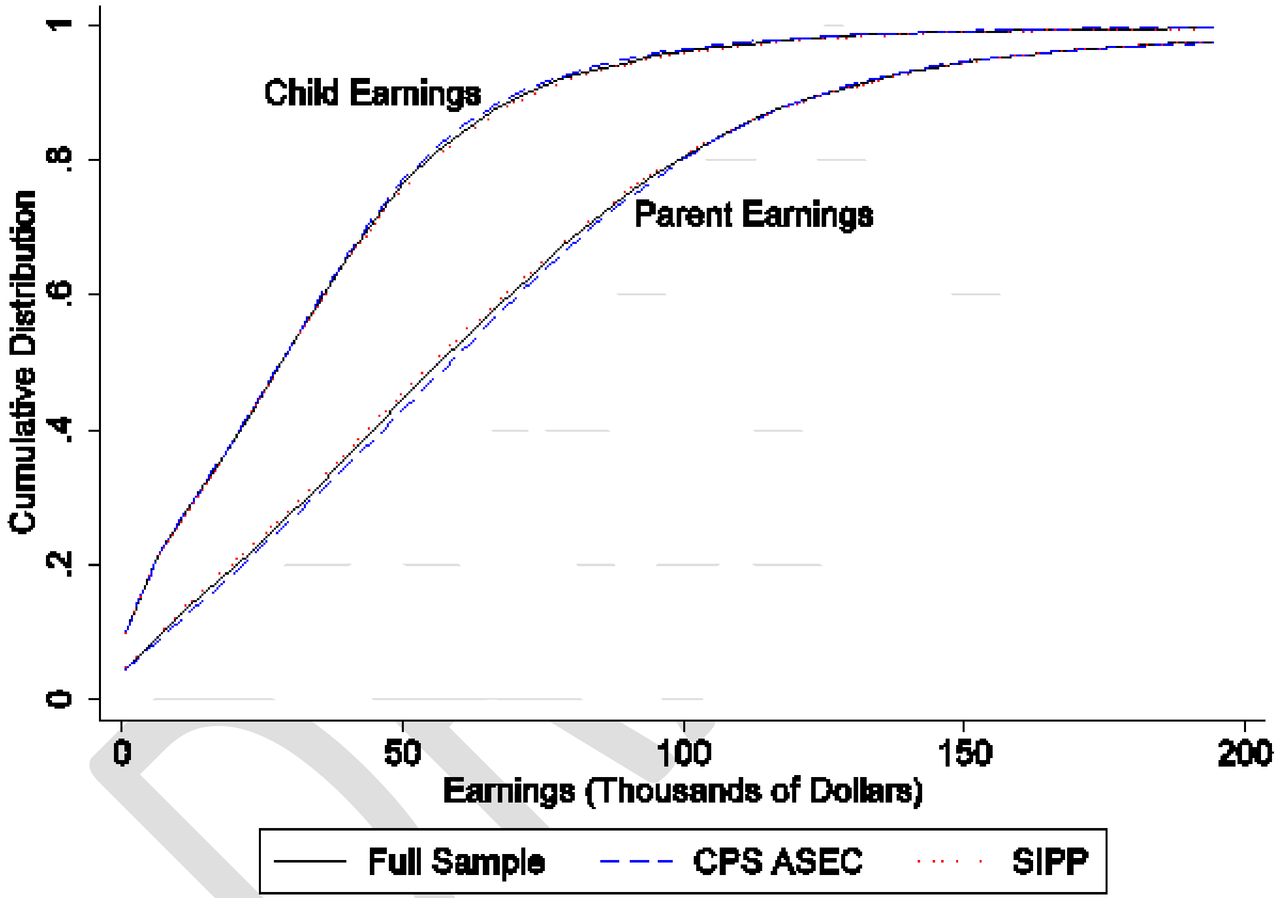
This figure plots the results of quantile regressions of child log earnings on parent log earnings and each of the family and demographic characteristics using cohort weights. The 95% confidence intervals are shown. This figure shows results from the 20th percentile as zero child earnings (recoded as 1) determine the coefficients at lower conditional percentiles. The full results are shown in Appendix Figure 5.

Figure 18: Unconditional Log Earnings Quantile Regression on Family Characteristics



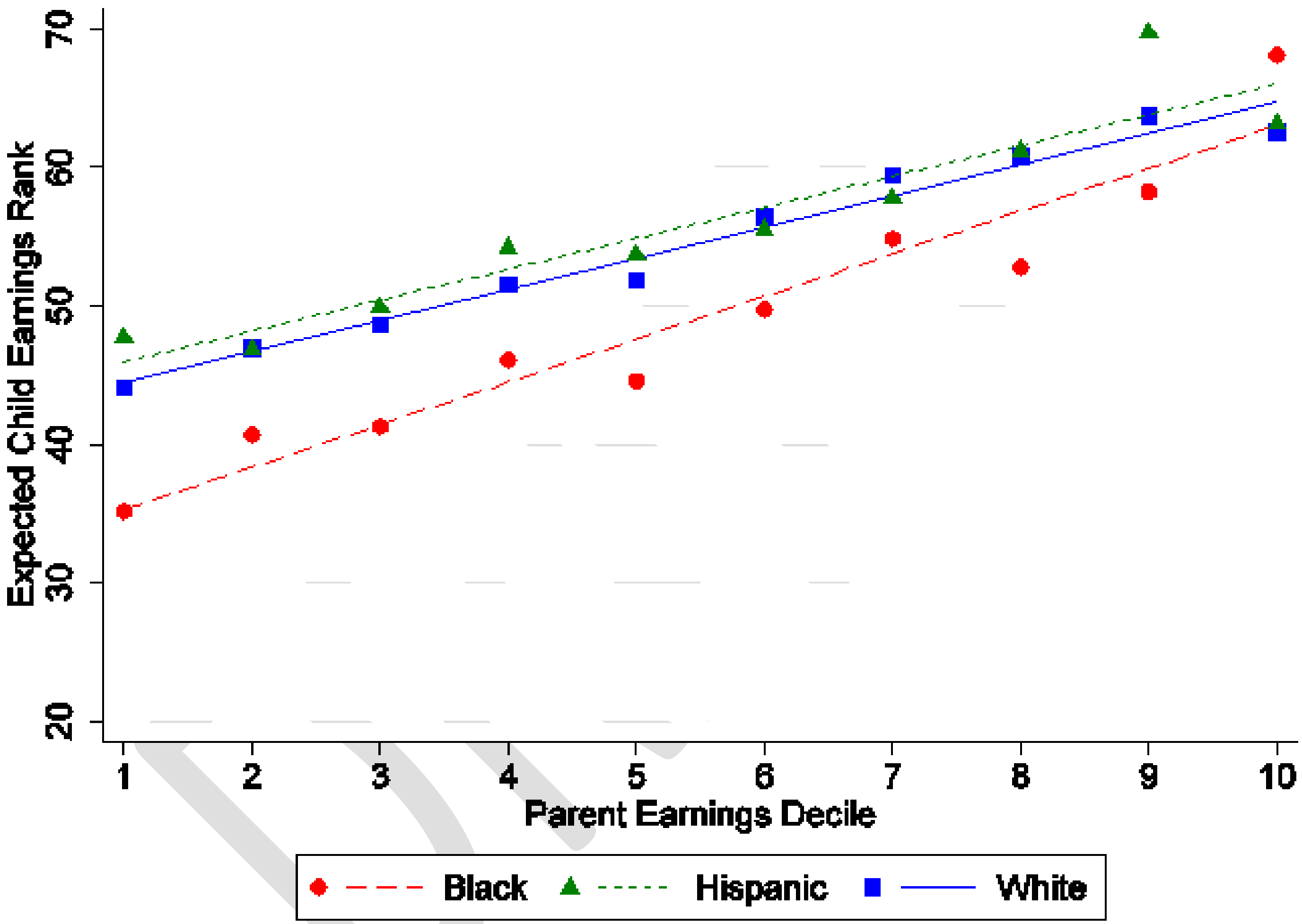
This figure plots the results of unconditional quantile regressions of child log earnings on parent log earnings and each of the family and demographic characteristics using cohort weights. This shows how the sample child distribution would differ if each characteristic where changed from the baseline (white, high school educated, married non-teen parents) to the characteristic in each subplot. This shows where in the distribution the lesser or greater of upward mobility of different subgroups has an effect. The 95% confidence intervals are shown. This figure shows results from the 20th percentile as zero child earnings (recoded as 1) determine the coefficients at lower unconditional percentiles. The full results are shown in Appendix Figure 6.

Appendix Figure 1: Parent Family and Child Individual Earnings Distribution in the CPS-SIPP/DER



The parent family earnings and child individual DER earnings are plotted for the CPS ASEC, SIPP, and combined full sample. Each line plots kernel density of the earnings in the relevant sample (in 2012 dollars). Parent earnings are much higher than child earnings for at least two reasons. First, parent earnings include the earnings of both spouses or partners whereas child earnings are for the individual children (as martial and partner data is not available for the children). Second, parent earnings are from later in their lifecycle as they are averaged when the older parent is between 40-44 years old, whereas child earnings are calculated when the children are 29-30.

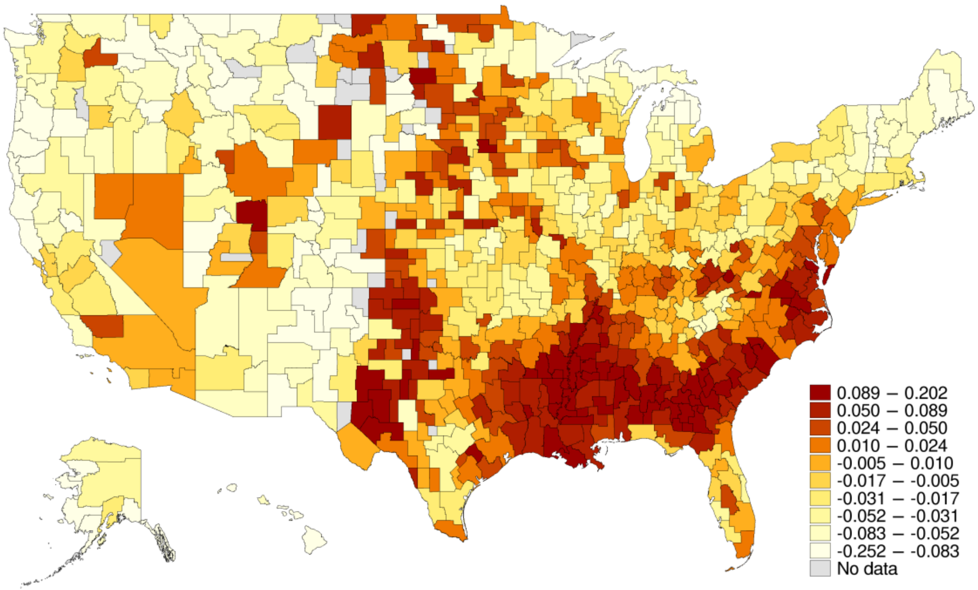
Appendix Figure 2: Race and Mobility by Decile Controlling for Parent Education



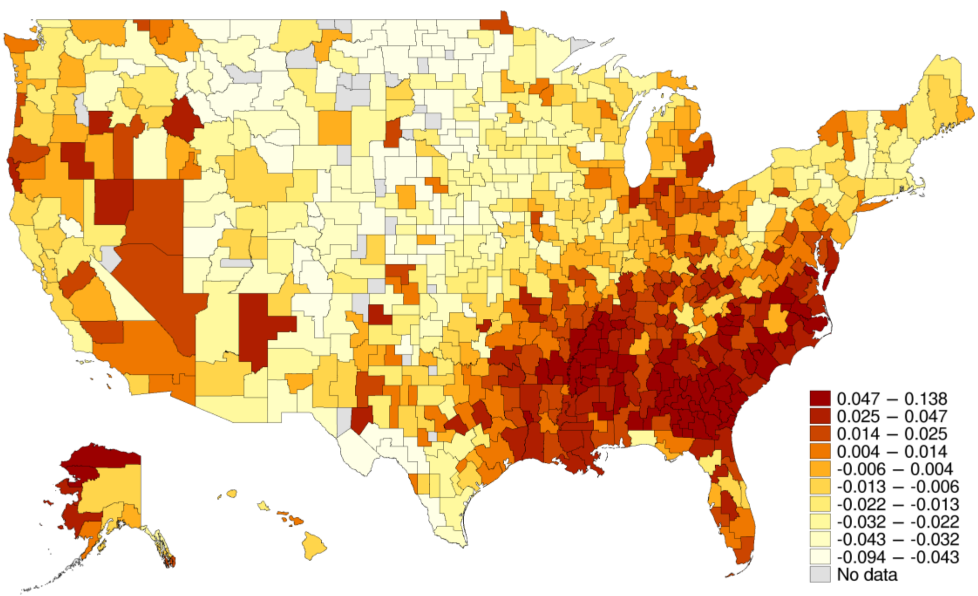
This figure plots the results of an OLS regression with dummies for each parent earnings decile interacted with race and highest parent education level (less than high school, some college, and college and above with high school as the default category). The three categories plotted by decile are white (and other), black, and Hispanic.

Appendix Figure 3: Impact of Family Characteristic Adjustment on Causal Mobility Estimates using Chetty and Hendren Permanents Residents at 26

A. 25th Percentile

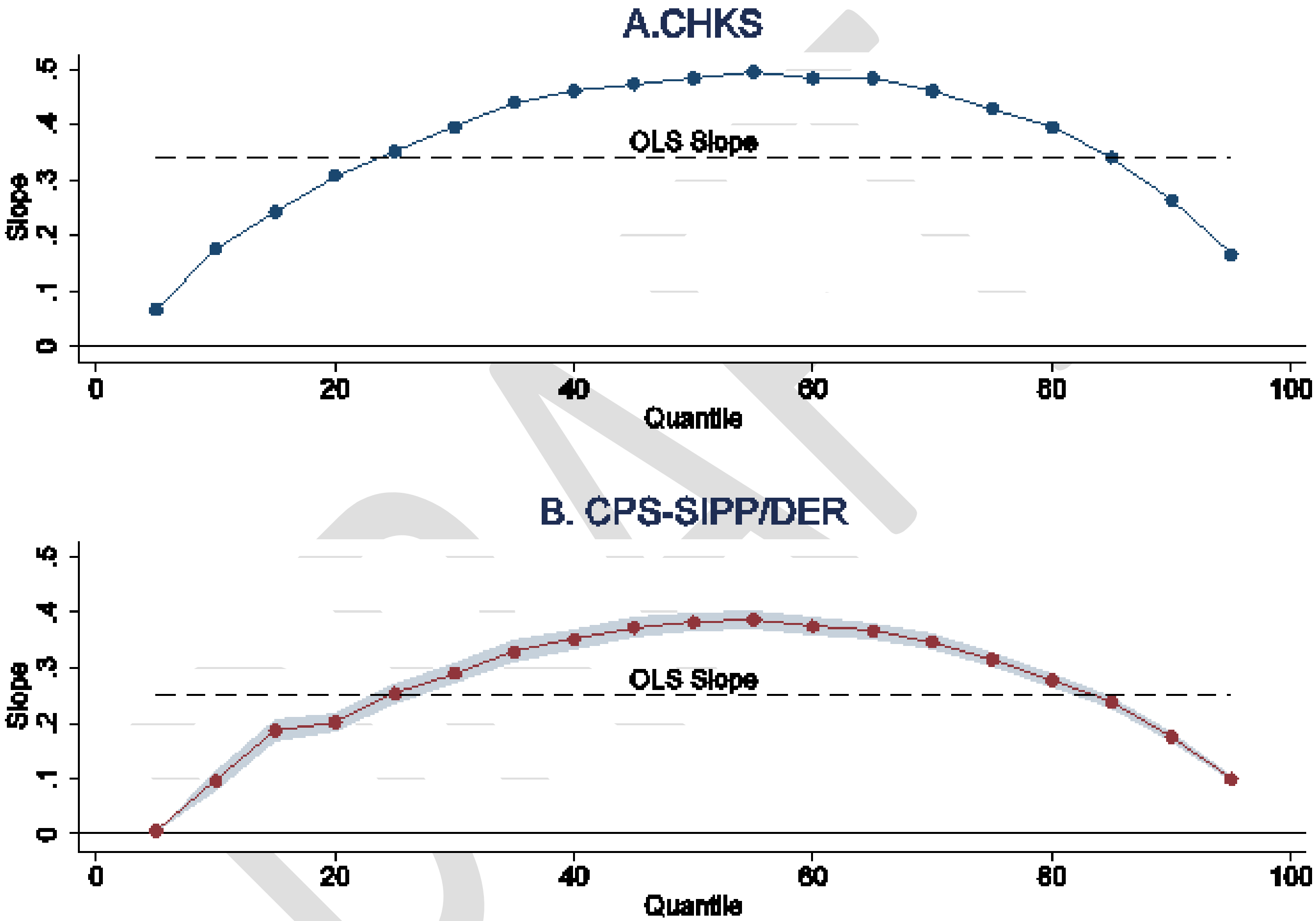


B. 75th Percentile



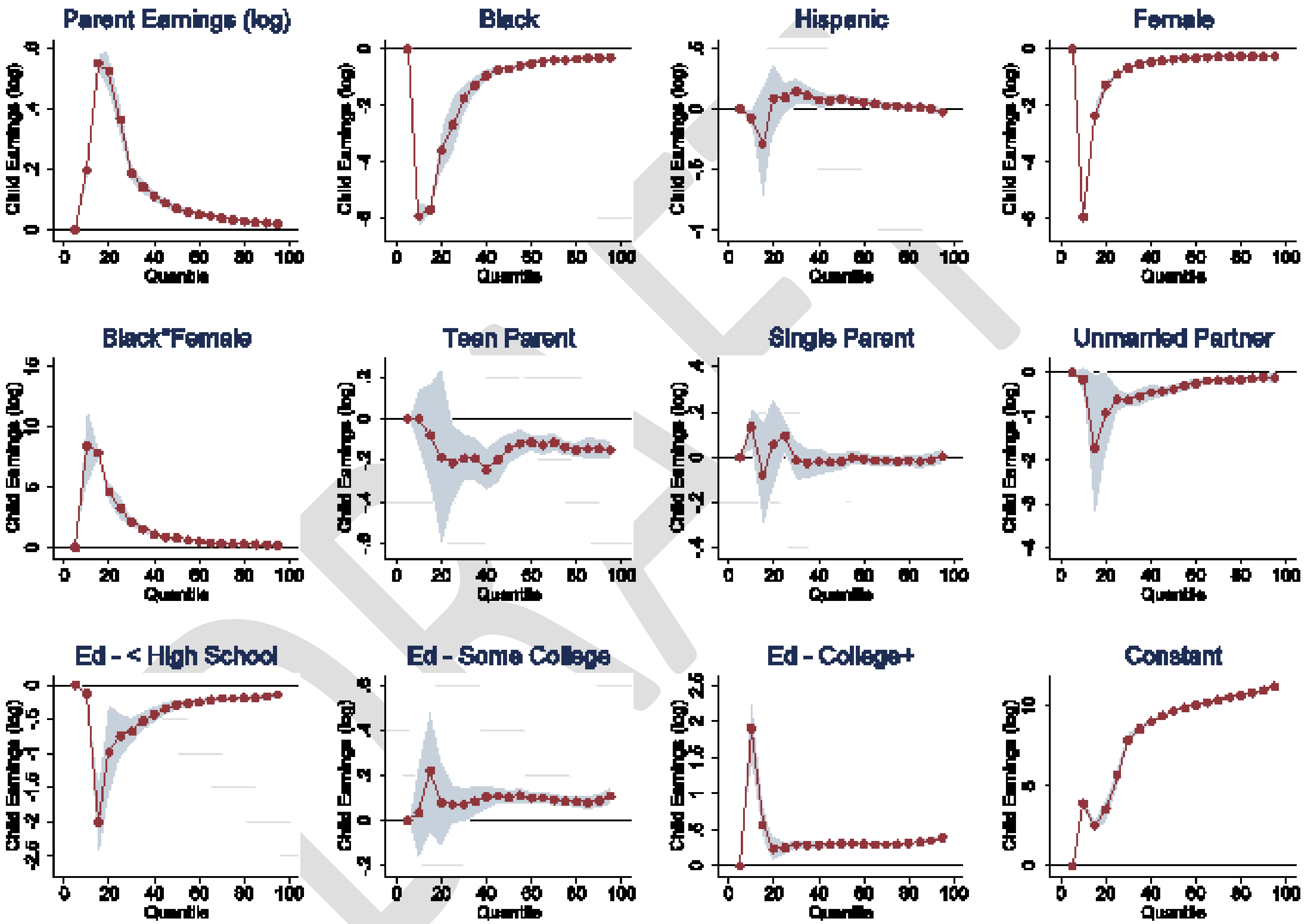
This figure shows how the forecast causal estimates from Chetty and Hendren is affected by replacing their estimates for non-mover mobility with the demographic and family characteristic adjusted estimates. The adjustment is calculated using the baseline model coefficients model (1) in Table 5 and using microdata from the 1990 census long form. For each child, the parent rank was estimated from the national distribution of parents in the 1990 census with older parents in the same age cohort and the child was assigned an expected child rank ($\hat{y}_i$) from the CHKS CZ-level slope and intercept terms. From the x characteritics in the baseline model observed in the census, I calculated $\hat{y}_{i,Adjusted} = \hat{y}_i - \beta_h h_i - \beta_{xh} x_i h_i$ and conducted the rank-rank regression of the adjusted child rank on the estimated parent rank, adjusting each CZ to the model baseline group of white sons of married, high school educated parents. Both maps are divided into ten equally sized quantiles. This figure differs from Figure 10 in the construction of the adjustment. In Figure 10, the adjustment is based on the CHKS estimates of mobility in each CZ (as that was the data used to construct the adjustment). In this figure, the adjustment is made directly to the permanent residents at 26 as in Chetty and Hendren. In each case, the adjustment to the permanent residents is the same, but the initial permanent resident value and forecast regression coefficient differ.

Appendix Figure 4: Conditional Quantile Regression of Child Rank on Parent Rank



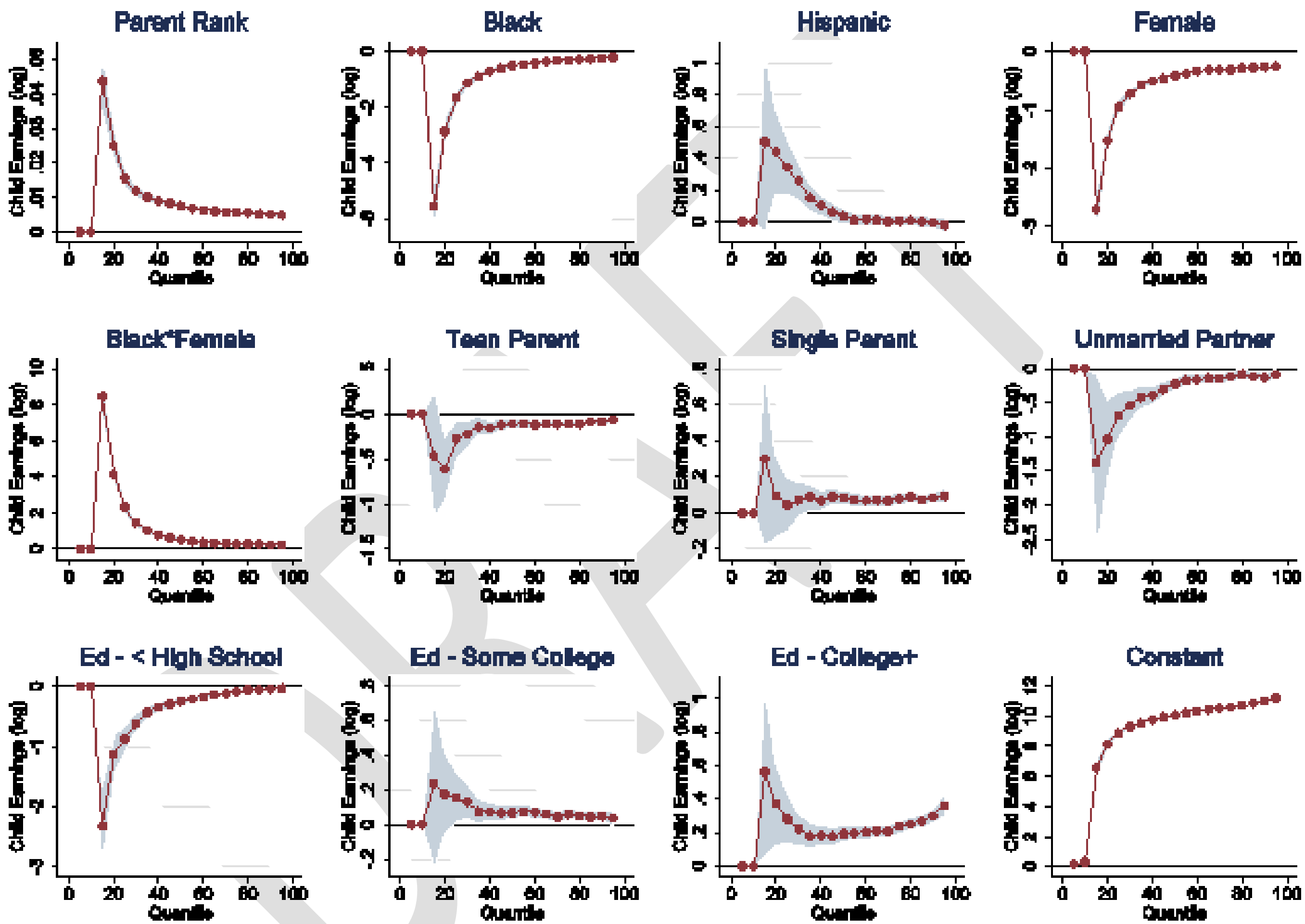
This figure shows the quantile regression of child rank on parent rank. Panel A shows the CHKS results (author's calculation from transition matrix data available at http://www.equality-of-opportunity.org/), and Panel B shows the results using the cohort-weighted CPS-SIPP/DER data with 95% confidence interval.

Appendix Figure 5: Conditional Log Earnings Quantile Regression on Family Characteristics



This figure plots the results of quantile regressions of child log earnings on parent log earnings and each of the family and demographic characteristics using cohort weights. The 95% confidence intervals are shown.

Appendix Figure 6: Unconditional Log Earnings Quantile Regression on Family Characteristics



This figure plots the results of unconditional quantile regressions of child log earnings on parent log earnings and each of the family and demographic characteristics using cohort weights. This shows how the sample child distribution would differ if each characteristic where changed from the baseline (white, high school educated, married non-teen parents) to the characteristic in each subplot. This shows where in the distribution the lesser or greater of upward mobility of different subgroups has an effect. The 95% confidence intervals are shown.

Appendix A

A. Data Construction and Weighting

This paper uses survey data to construct a large sample of parents and children linked with administrative longitudinal earnings data. The parent-child links and information on parent family characteristics come from two surveys conducted by the US Census Bureau, the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the Survey of Income and Program Participation (SIPP). The earnings data comes from W-2 earnings records filed by employers with the Social Security Administration and the Internal Revenue Service and shared with the Census Bureau in the Detailed Earnings Record (DER) extract from the SSA Master Earnings File. Individuals are linked between the Census surveys and the DER by matching survey respondents to their Social Security Numbers (SSN).[27] Prior to the construction of the Census survey-administrative data set, the SSNs are removed from the data and individuals are given a Personal Identification Key (PIK) to enable the linkage.

While the CPS ASEC has been conducted annually since 1948, the links between the SSNs and respondents are currently available for the following survey years: 1991, 1994, 1996-present. The data in this paper uses the linked CPS ASEC files up to 2009. The SIPP data used in this study comes from an internal data product at the US Census Bureau, the SIPP Gold Standard File (GSF). It contains all SIPP respondents from the 1984, 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008 panels. However, in this paper, I do not include observations from the 1984 panel of the SIPP as the family relationships were not gathered until the Wave 8 topical module conducted from January to March of 1986 and are therefore not available for families that attritted out of the sample. In the CPS ASEC only children aged 15 and older were given a PIK to allow matching to the DER, and I only include children observed in their parent household up to age 18 in my sample. The DER earnings file contains annual W-2 earnings information from 1978 to 2012.

[27] The process by which CPS ASEC and SIPP individuals are linked to the DER file is described in Wagner and Layne (2014).

A.1 Weights

The CPS ASEC and SIPP both provide weights for individual observations in each round of the survey based on their probability of selection and response. However, as I am combining parent-child pairs over two dimensions: 1) across multiple survey rounds for the same survey and 2) between the two surveys, I have chosen to adjust the within survey-year weights to more accurately reflect the child population.

To weight observations across multiple survey rounds, I group children by age cohort. For example, a child who is 16 in the 1994 CPS ASEC would be in the 1978 cohort, as would a child who is 15 in the 1993 SIPP panel. Because the number of parent-child pairs varies by child age cohort, I normalize across cohorts so that the sum of the weights is one for each child age cohort.

This normalization is done for the CPS ASEC and SIPP samples separately before combining the samples. To combine the two samples, I adjust the weights by the share of the total number of observations for a given child age cohort that comes from that survey. So if share $\alpha \in [0,1]$ (of the unweighted number of observations) of the 1978 cohort comes from the SIPP and $1-\alpha$ from the CPS ASEC, then the SIPP observation weights are multiplied by α (and sum to α) and the CPS ASEC weights by $1-\alpha$ so that the sum of the weights for the combined sample is again 1 for the child age cohort. In this way, the average weight of an observation is the same whether it comes from the CPS ASEC or SIPP sample.

To be included in the CSD sample, each parent-child pair must be matched to their SSNs. A pair is successfully matched if the child and all parents (both parents in two-parent families and the individual parent in one-parent families) are successfully matched. The match rates for the CPS ASEC and SIPP samples by child age cohort is reported in Table 1. For all cohort groups, the average match rate across the two surveys is above 70%.

A.2 Ages of Earnings Observation

The next step is to determine at what ages to measure parent and child earnings for the intergenerational mobility comparison. For parents, I average family earnings over the 5 years when the older parent is 40-44 years old. This was chosen for two reasons. First, the literature on life-cycle bias in estimates of intergenerational mobility suggests measuring income around 40

(Haider and Solon 2006). Second, this choice allows me to better compare my results to CHKS as they use a 5-year average of parent income in their analysis.

For children, the issue is complicated by sample size concerns. Because the earliest available surveys that can be matched to the DER are from 1991 for the CPS ASEC and 1990 for the SIPP, there is a tradeoff between observing children at later ages and reducing the sample size. For example, the oldest possible child in my sample is 18 years old in the 1990 SIPP. This child would be 39 in 2011, the final year of the available DER earnings data. However, if I restrict my sample to only those who are 39 by 2011, my sample would include only 533 parent-child pairs.[28] Instead, I follow CHKS in focusing on children around the age of 30. They show that there is little lifecycle bias in rank-rank income mobility by age 30 in child income. To test for lifecycle bias in the CSD sample, I plot the rank-rank slope of intergenerational earnings mobility with child earnings measured over two years starting from age 24 to 32, shown in Figure 1.[29] Panel A shows the effect of measuring earnings by age for the full sample. The general trend is similar to that in CHKS with increases at younger ages and potentially slight decreases at higher ages, but few of the differences are statistically significant. Panel B shows the trend for male children, which is increasing up to about 29 and flat above. The slight downward trend in Panel A is due to a decrease in the rank-rank slope for female children.

I have chosen to use average child earnings at 29 and 30 for the baseline sample to more closely match the period used in CHKS, where income was measured starting at 29-32 years old depending on the child's age cohort and to maximize the sample size.

A.3 Assigning Parent and Child Ranks

In order to proceed, I must assign ranks to each parent family and child individual earnings level. The sample comes from a wide variety of parent and child age cohorts. If I use earnings from the same calendar years, then I am comparing individuals at different stages in their life cycle. However, if I use earnings at the same age, then the comparisons will be over vastly different stages in the business cycle or even as far apart as three decades. For example, there are parents in my baseline sample who turn 40 in 1978 and others who turn 40 in 2007. Instead, I have chosen to compare parents and children to samples of all matched parents and children in their

[28] This includes the restriction that the older parent turns 40 between 1978 and 2007.
[29] The method for converting earnings to ranks is discussed in Appendix A.

age cohort in the CPS ASEC and SIPP. The parent that turns 40 in 1978 would be compared to all parents that turn 40 in 1978 regardless of the age of their children or the survey and year in which they were observed.[30]

To construct the parent comparison groups, I create a sample of all parent families from any year in either survey where all parents have PIKs, indicating a successful match to the SSN database. To be in this comparison group, a child must be present in the household, but the child need not be matched (either because a match was not available or because the child was 14 or under in the CPS ASEC and no match was attempted).

For the child comparison sample, I make a simplifying assumption which vastly increases the size of the comparison group. I include all adults in the child's age cohort observed in any survey year as part of the comparison group, thereby assuming that in- and out-migration are sufficiently small between the year the child was observed in the CPS ASEC and SIPP and the year the adult cohort was observed in the later survey. In this way, a child born in 1980 and observed at 16 in the 1996 CPS ASEC would be compared to all matched individuals born in 1980 from either survey, including a 29 year-old adult observed in the 2009 CPS ASEC or a 24 year-old adult observed in the 2004 SIPP.

For the baseline sample of children at 29-30 and parents at 40-44, the earnings distributions are shown in Appendix Figure 1 for the full CSD sample and separately for the CPS ASEC and SIPP subsamples. Because the parent earnings are for families and the child earnings are for individuals as well as due to the later age of parent observation, parents earn much more than children in the sample. There are also more children than parents with zero earnings (as in CHKS with income).

[30] For two parent households, I use the age of the older parent.