Parent Characteristics and the Geography of Mobility

Jonathan Rothbaum*

March 23, 2015

Abstract

There is tremendous spatial variation in intergenerational upward mobility in the United States. Understanding the sources of this variation is important to understanding how to foster upward mobility and equality of opportunity. In this paper, I use data from Census surveys linked to administrative tax data to explore to what extent spatial variation in parent family characteristics and demographics can account for the observed pattern of spatial variation in mobility. Parent education and child race are highly predictive of child earnings rank even after controlling for parent earnings, and these characteristics are also highly correlated with upward mobility. However, the results in this paper suggest that none of the lower earnings of Black children or the higher earnings of children of more educated parents are due to the places in which they live. An important share, but by no means all, of the spatial variation in mobility can be accounted for by the variation in child race, parent education, and parent family type (single or teen parenthood). In addition, many of the local area characteristics that are highly correlated with mobility, such as the share of single parents, income inequality, and social capital, are not predictive of mobility after controlling for these family and demographic characteristics. As an additional step to understanding the variation in child outcomes, I explore the heterogeneity in the relationship between parent characteristic and child earnings using quantile regressions.

* Social and Economic Housing Statistics Division, U.S. Census Bureau. Email: jonathan.l.rothbaum@census.gov
This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any error or omissions are the sole responsibility of the author. Any views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau. All data used in this paper are confidential. Calculated standard errors do not account for the CPS ASEC or SIPP sample designs. The estimates in this paper are based on responses from a sample of the population. As with all surveys, estimates may vary from the actual values because of sampling variation and other factors. All comparisons made in this paper have undergone statistical testing and are significant at the 95 percent confidence level unless otherwise noted. Estimates are calculated from the portion of SIPP and CPS ASEC sample matched to administrative records, and so are not fully representative of the national population.

1 Introduction

The existence of 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.[1] 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, Chetty et al. (2014), hereafter CHKS, are able to characterize the geographic variation at a much finer level. They find that parent income is a much stronger determinant of child income in much of the Southeast 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.

The next logical question is: what characteristics of individuals or local areas are responsible for the variation in upward mobility? Understanding this has enormous implications for understanding intergenerational mobility and fostering equality of opportunity in the United States.

As a first step in the analysis of potential causes of this heterogeneity within the United States, CHKS test for the relationship between a variety of local area characteristics and local 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

[1] Denmark, Finland, Norway, and Sweden

regressions. On the other hand, several long-term longitudinal surveys 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). The PSID and NLSYs have been studied in the context of mobility 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.[2] This limits the types of 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).

In this paper, I make use of administrative tax data linked to two Census surveys, the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) and the Survey of Income and Program Participation (SIPP). The surveys provide the link between parents and children and information on the characteristics of the parent families. The administrative 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. While the resulting sample is much smaller than the one used by CHKS, it is also much larger than the samples in the PSID and NLSYs.[3]

As a first step, I replicate the main results in CHKS to show that they hold for intergenerational mobility of earnings as well as of income. For example, I show that there is a high correlation between the geographic variation in mobility using W-2 earnings data and the variation found in CHKS with income data. I then add a small set of family and demographic characteristics, including race and parent education, to their baseline rank-rank mobility regressions and show they account for an important share, but by no means all, of the spatial variation in mobility. I also use their results to test whether omitting spatial variation from

[2] 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.

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

regressions of child earnings on parent characteristics biases results and find little evidence that it does.

I confirm that many of the relationships found by CHKS between local area characteristics, such as inequality and the rate of single parenthood, and upward mobility are also present in the CPS-SIPP/DER data (CSD). However, after including the family and demographic characteristics, none of these relationships remain statistically significant in the CSD data.

The literature is also rich in analyzing the relationship between parent characteristics and child outcomes. This includes a large number of papers, including 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.

I contribute to this literature by documenting the considerable heterogeneity in mobility using quantile regressions and show how this heterogeneity is related to parent characteristics. For example, I show that race or having a teen parent is especially strongly associated with lower earnings for otherwise high achieving children but has a smaller impact on otherwise low achieving children. I also use unconditional quantile regressions to show where in the child earnings distribution these mobility effects are concentrated.

As with much of the literature on mobility, these relationships should not be interpreted as causal. For example, I show a very strong association between parent education and child earnings, even after controlling for parent earnings. However the causal estimates on the relationship between parent and child outcomes in the literature are ambiguous. Using changes in compulsory schooling laws, Black, Devereux, and Salvanes (2005) find limited evidence of a causal relationship on child education in Norway, whereas Oreopoulos, Page, and Stevens (2006) find a strong causal relationship between parent education and the probability that a child repeats a grade in the United States.

The paper is organized as follows. In section 2, I describe the CSD 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 on heterogeneity related to parent characteristics. Section 6 concludes.

2 Data

This paper uses data from several surveys 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 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).[4] 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 2011.

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, the CPS-SIPP/DER (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

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

as well. The most obvious is that the CSD sample is orders of magnitude smaller than the CHKS administrative tax records. In my baseline sample, there are 41,990 parent-child pairs compared to 9,867,736 in theirs. In addition, the DER file only contains information about wage and self-employment earnings taxable by Social Security, but not about the other income sources included in CHKS. Finally, the 1040 data available to CHKS also contains information on the marital status and income of the children's spouses. This information is not present in the DER. This limits me to analyses of intergenerational earnings mobility between parent families and their individual children.

2.1 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.[5]

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 is in the same age cohort as a child who is 15 in the 1993 SIPP panel and another who is 17 in the 1995 CPS ASEC. 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 2010 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

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

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%.

2.2 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.[6] 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.[7] 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. 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.

[6] This includes the restriction that the older parent turns 40 between 1978 and 2007.
[7] The method for converting earnings to ranks is discussed in Section 2.4.

2.3 Summary Statistics

In Table 2, for my baseline sample (parents at age 40-44 for the oldest parent and children at 29-30), I report summary statistics for family and demographic characteristics 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. This means that if a child was first observed in the 1996 CPS ASEC living with one parent, the child is categorized as being from a single-parent family even if in 1997, two parents were present (when the child may or may not have been observed in the survey). The summary statistics are calculated using the cohort weighting method discussed above.

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.[8] 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 highly educated parents.

Despite this selection, the matched sample is broadly representative of the underlying full set. The matched sample is 13.0% Black, compared to 14.4% in the full sample. For Hispanics, the share in the matched sample is 9.7% compared to 11.0% in the full. In the full sample, 13.3% of children have parents with less than a high school education, 33.6% with a high school education, 27.7% with some college, and 25.4% with at least a college degree. In the matched sample, the corresponding shares are 11.6%, 33.1%, 28.8%, and 26.6%. All differences except for parents with high school education are statistically significant.

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

[8] 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).

option on either survey in the years studied.[9] The estimated partner families are not statistically significantly different between the full and matched sample with 2.5% partner families in the full and 2.2% in the matched sample.

2.4 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 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.

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.

[9] 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.

Even in age cohorts with relative few parents or children in the matched parent-child pairs, I can estimate the rank of each from much larger comparison groups. This also helps reduce the bias from estimating the rank from non-representative samples. For example, parents who turn 40 in 2007 with children that are old enough to observe their adult earnings in my baseline sample (29 in 2009) are by definition teen parents. To compare those parents to only other parents in their age cohort with children that are old enough to be observed as adults would likely bias their rank upwards. By comparing them to all matched parents in their age cohort whether their children are old enough or not to be in the CSD sample, this bias can reduced or eliminated altogether.

For the baseline sample of children at 29-30 and parents at 40-44, the earnings distributions are shown in Figure 2 for the full CSD sample and separately for the CPS ASEC and SIPP samples. 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).

3 Intergenerational Mobility of Earnings Rank

As a first step in the analysis of intergenerational mobility of earnings rank, I attempt to replicate to the extent possible the results in CHKS. In doing so, I hope to show that their results hold despite the differences between the data sets used. By replicating their results, I will be able to build on their work using the family and demographic information available in the Census surveys with more confidence that any findings are due to the additional family information and not other differences in the income and earnings data or parent-child samples.

As noted previously, the first difference between the two data sets is that CHKS use data on a larger variety of income sources.[10] In the CSD sample, I have only wage and self-employment earnings from W-2 filings. The second important difference is that CHKS can infer the marital status from tax filing information 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

[10] CHKS 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).

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 my baseline sample.

3.1 Rank-Rank Mobility

CHKS estimate the rank-rank mobility for a set of income definitions, including parent family income rank→child family income rank, parent family income rank→child individual earnings rank, 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. Unfortunately, they do not report any coefficient for mobility of parent family earnings rank→child individual earnings rank, which would correspond more closely to what can be estimated using the CSD data. I therefore compare my results to the closest analogue in CHKS, parent family income→child individual earnings.

One advantage of estimating mobility as the rank-rank slope of parent and child income is that the relationship is linear, which CHKS show is not true for log income. Another advantage of measuring mobility of rank is that the inclusion and treatment of zeroes is straightforward, whereas with log income and the intergenerational elasticity of income, the coefficient is highly sensitive to these decisions. Figure 3 shows a binned scatter plot of average child rank and parent rank from CHKS and the CSD samples. 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.

In Table 3, I report the rank-rank regression coefficients estimated by CHKS and using the baseline CSD sample using cohort weights, including for male and female children separately. The CHKS coefficient for parent family income→child individual earnings is 0.282 compared to 0.245 for parent family earnings→individual child earnings in the CSD sample, a difference of 0.037 (or 13.1%). 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.[11]

[11] Although only at the 10% level for females in the CSD sample.

In their paper, CHKS divide the United States into 741 commuting zones (CZ).[12] For each commuting zone (c) with at least 250 children in their sample, they estimated the slope (α_c) and intercept (β_c) of the parent-child rank-rank mobility regression using their baseline sample and income definitions.[13] 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.[14]

3.2 Geographic Heterogeneity

From this analysis, it is clear that there are differences between the rank-rank slope found in CHKS and in 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 paper is the geographic heterogeneity of mobility.

In this section, I will show that this heterogeneity is also present in the earnings data. I do so in two steps. First, I estimate the same CZ-level rank-rank regression models as in CHKS using the CSD data. The results suggest that the correlation in CZ-level mobility in the two data sets is high. However, the CSD sample size is not sufficiently large to draw a stronger conclusion. Second, I group CZs based on their level of mobility as estimated by CHKS. This allows me to get more precise estimates on the mobility estimates in the CSD sample. I show that there is a very high correlation between the CHKS and CSD mobility estimates for these CZ groups.

In the aforementioned first step, I estimate the correlation between the CZ-level rank-rank regression coefficients in CHKS and with the CSD data using the regression:

[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 United States (continental, Alaska, and Hawaii).
[13] CHKS data on CZ mobility coefficients is available at http://www.equality-of-opportunity.org/.
[14] The coefficient of 0.700 is very close to the ratio of the rank-rank coefficients for the CSD and the CHKS baseline of 0.341, where $\frac{0.245}{0.341} = 0.718$.

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

For CZ c, the CSD coefficients are denoted as α_c^{CSD} and β_c^{CSD} and for CHKS as α_c^{CHKS} and β_c^{CHKS}.[15] The correlations I calculate are between 1) the intercepts α_c^{CHKS} and α_c^{CSD} and 2) the slopes β_c^{CHKS} and β_c^{CSD}.

However, given the CSD sample of 41,990 parent-child pairs across 570 CZs, there are only 74 children per CZ on average. CHKS only includes CZs in their analysis with at least 250 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 intercept and slope correlations for the included areas. At one extreme, including only CZs with at least 1,000 observations leaves only three in the analysis and unweighted correlation coefficients of 1.00 (0.9985) and 0.99 for the intercept and slope respectively. At the other extreme, including CZs with three observations leaves 528 and unweighted correlations of 0.22 and 0.11 for the intercept and slope respectively. At the CHKS minimum of 250 observations, there are 32 CZs in the CSD sample, all larger metropolitan areas, with unweighted correlations of 0.69 and 0.76 for the intercept and slope.[16]

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 without the problem of small size that is apparent in Table 4 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 the CZ-level rank-rank slope coefficient in 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, and the 50th group contains the 14 with the highest slopes.[17] For each group, I estimate the benchmark CHKS slope (β_k^{CHKS}) and intercept (α_k^{CHKS}) by averaging the CHKS CZ values across the individual CSD

[15] As with CHKS, all regressions are run using the parent and child ranks in the national distribution earnings, not estimated from each CZ distribution.

[16] 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.

[17] With 50 groups, there are approximately 14 CZs per group as $\frac{709}{50} \approx 14$.

observations.[18] The observation-weighted CHKS intercept and slope are plotted in Figure 4 Panel A, where k =5, 10, 20, 25, and 50 (and $p = 1, \ldots, k$ for the regression in each group p of $y_{ip} = \alpha_p + \beta_p x_{ip} + e_{ip}$).

I estimated the slope (β_k^{CSD}) and intercept (α_k^{CSD}) coefficients in the CSD data by regressing child earnings rank on parent earnings rank in each group. 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 slope and intercept coefficients. The results are shown in Figure 4 Panel B. In the CSD sample, the estimated slopes are lower in magnitude and the intercepts greater in magnitude than CHKS, which reflects differences between the CHKS income data (slope of 0.341) and the CSD earnings data (slope of 0.245). Although there is more variation in the CSD data around the trend, the pattern of spatial heterogeneity found in by CHKS also exists in the CSD data.

I calculated 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 142 CZs per group), the smallest group has nearly 5,600 observations and the correlation between both the slope and intercept in the CHKS and CSD is almost perfect (0.99 for both). At 25 groups (28 CZs per group), there are at least 547 observations in each group, and the correlations are 0.94 for the intercept and 0.86 for the slope. At 50 groups, there are at least 162 observations in each group[19] and the correlations are 0.87 for the intercept and 0.72 for the slope.

Taken together, the results in Figure 4 and Table 4 indicate that the there is a high correlation between the spatial heterogeneity found in 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.

[18] 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 $\frac{0.39+0.36+0.30}{3} = 0.35$.

[19] Two groups have fewer than 250 observations, 15 have fewer than 500, and 41 have fewer than 1,000.

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 ran a simple semiparametric model using 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)

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

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 and standard errors are plotted in Figure 5 for Whites, Blacks and Hispanics.[20] There are no statistically significant difference between the outcomes of White and Hispanic children holding the parent and family characteristics constant. Blacks at lower and middle income deciles have lower expected ranks than Whites (1^{st}-8^{th}) and Hispanics (1^{st}-6^{th}), but the outcomes of Black children converge to those of Whites and Hispanics at higher parent earnings. However, the differences at lower parent ranks are very large in economic terms as the average rank difference between Black children and White and Hispanic children where they are statistically significant is 8.3 percentile. This is similar to findings by Bhattacharya and Mazumder (2011) in their analysis of upward mobility differences between Blacks and Whites using NLSY data. The gap between Black and White upward mobility has also been found in PSID data by Hertz (2005) and in a SIPP/DER sample (and NLSY) by Mazumder (2011).

Figure 5 suggests that the relationship between parent rank and child rank is roughly 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 (not shown). As a result, I proceed using a simpler parametric model of parent characteristics and parent and child outcomes:

[20] 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. The differences hold for other groups (college graduates, teen parents, etc.) with a shift of each line by the same constant at each decile. The standard errors are also affected by choosing a different value for the category dummies depending on the size of that group at each decile. A more detailed set of interactions between race and education groups at each decile was tested but the standard errors are too wide to be useful due to the small sample sizes in many race-parent education-parent earnings decile cells. The errors for all regressions are clustered at the CZ level.

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

In Table 5, 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 above, although the results are largely unaffected by the use of weights. In column (1), I include race dummies and race dummies interacted with parent rank. With race dummies and interactions, the coefficient on Black is negative and statistically significant. In addition, the coefficient on the Black-parent rank interaction is positive and statistically significant. In column (2), I include the other family characteristics and interactions between parent rank and gender and parent education. The coefficient on all of the uninteracted variables are significant. I have chosen the model in column (3) as the baseline. This model includes the interactions of Black, Hispanic, and female with parent rank 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, and for both, the point estimates are large in magnitude.

In the baseline model in column (3), there are a number of notable results. The coefficients on Black have decreased in magnitude (although not statistically significantly) from column (1) and both the dummy (intercept) and dummy interacted with rank (slope) are still statistically significantly different than for White children. The greater slope for Black children means that parent earnings are a more important determinant of Black child outcomes. For Black children, upward mobility is lower than for White or Hispanic children controlling for parent education, marital and cohabiting partner status, and teen parenthood. The magnitude of the difference is worth noting. Given a White child whose parents' earnings are at the median, for a Black child to have the same expected rank, his parents would have to be above the 70$^{\text{th}}$ percentile. For White parents at the 25$^{\text{th}}$ percentile, the Black parents would need to be above the 51$^{\text{st}}$ percentile for their children to have the same expected rank.

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 35 percentile higher in the parent rank distribution ($\frac{7.80}{0.221} = 35.3$). 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 ($\frac{7.02}{0.221} = 31.8$). In both cases, the point estimates on the interaction terms suggest this gap narrows, but only slightly, 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 0.245 to 0.221, or 9.8%. Parent education is an important predictor for child earnings rank 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 on the baseline model in Table 5 column (3) could be biased. For example, if Black children 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 White 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.

To test whether spatial heterogeneity is biasing the baseline results, I replace the parent rank and all interactions of rank in the baseline model with the CHKS CZ-based predicted child rank (from equation (3.2)). 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 children, the coefficient on Black in the baseline predicted rank regression should be lower than in the parent rank regression.

The results are shown in Table 6. In column (1), the baseline model results on parent rank are shown, and in column (2), the results are shown using CZ-level predicted 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.[21]

What has changed is that the spatially adjusted slope coefficient for Blacks is no longer statistically significantly different from Whites, although it somewhat large in magnitude, 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, the result implies that better performance of Black children from high income parents may be because they are more likely to live in high mobility areas than Black children from low income parents. In other words, very little, if any, of the lower upward mobility experienced by Black children is due to the lower mobility of the places in which they live.

It is striking in comparing columns (1) and (2) just how little some of the coefficients on parent characteristics change. The level (uninteracted) coefficients 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 either unaffected by, or they are completely uncorrelated with, 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.

[21] 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 clearly in the interaction for female as the distribution of female children should be uncorrelated with spatial heterogeneity. The female slope adjustment changes from -0.083 in model (1) to -0.248 in model (2) with a ratio of $-\frac{2.48}{-0.083} = 2.99$

Next, I test the extent to which these parent characteristics are correlated with upward mobility in the CSD data. I do so in two ways. First, 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 (equation (3.3)).

The results are shown in Table 7 Panel A. At the CZ-group level, Black children, high school educated parents, teen parents, and single parents are highly correlated with low mobility. Parent college education (some or college and above) and potentially Hispanics are positively correlated with high mobility. To ensure that results do not depend on the CZ groupings in Panel A, I regress the CHKS intercept and slope on the family and demographic characteristics at the individual level, shown in Panel B. Column (1) shows the intercept regressed against each characteristic separately, and column (3) shows the same for the slope. In both cases nearly all of the characteristics are statistically significantly correlated with mobility except unmarried partners, parents with less than a high school education, and the relationships correspond to the results from Panel A. Columns (2) and (4) show the results of regressing the intercept and slope respectively on all of the characteristics simultaneously. In these columns, single parenthood is no longer a significant predictor of mobility after controlling for the other characteristics.

Although the spatial variation of many of these characteristics is correlated with the variation in mobility, accounting for the variation in mobility does not have a statistically significant (or economically significant) effect on the regression coefficients for many characteristics. For those coefficients that are impacted, the results, though far from conclusive, suggest a greater impact of parent characteristics on child outcomes controlling for spatial heterogeneity, not less.

This leaves open the possibility that spatial variation in these individual and family characteristics could explain a part of the variation in intergenerational mobility found by CHKS. I test for this by calculating an adjusted rank for each child that controls for the observable demographic and family characteristics. I create the 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 2 column (1), where

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

As before, I 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

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

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 some 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 8. The intercept regression coefficient (δ_α) is never above 0.58 for any of the quantile groups from 5-50. This implies that the observable characteristics may explain up to 42% of the variation in the intercept. The slope regression coefficient (δ_β) varies from 0.63 (5 groups) to 0.72 (50 groups), which implies that up to 28% of the variation could be explained by the observable characteristics.[22]

For the measures of dispersion, the observable characteristics reduce the variation of the intercept term by at least 38% across all measures and number of quantile groups (and as much as 71% for the variance with ten CZ-quantile groups). For the slope term (CHKS relative mobility), the reduction in dispersion is lower, but still at least 16% over all measures and group sizes (and up to 55% for the variance with five groups).

I also tested the same model omitting the interaction between Black and parent rank (as it is not significantly different from zero in the predicted rank regression in Table 6) and the results are nearly identical to the ones reported. The results in this section suggest that demographic and family characteristics can potentially explain an important share of the observed heterogeneity in upward mobility found by CHKS. In other words, where a child lives does influence their

[22] For both δ coefficients, the value is equal to the correlation times the ratio of the standard deviations of the adjusted and unadjusted α_p and β_p coefficients.

upward mobility, but not as much as CHKS estimated because they could not control for many parent characteristics.

4.2 CZ Characteristics and Upward Mobility

Beyond documenting the spatial heterogeneity in upward mobility, CHKS also analyze geographic correlates to that mobility. They identify five main factors at the CZ-level (with the measure they use in parenthesis): 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 by Putnam (1995)), and 5) rate of single parenthood (fraction of single mothers). As they stress, these relationships are not necessarily causal as there are many unobserved differences across areas that they do not control for in their analysis. Individual and family characteristics are one such set of variables that are not available in their data.

First, I test whether including the CZ-level characteristics (z_{ic}) affects my baseline results, by including all five factors in the baseline model

$$y_i = \beta_x x_i + \beta_h h_i + \beta_{xh} x_i h_i + \beta_c z_{ic} + \beta_{c,slope} x_i z_{ic} + e_i. \tag{4.3}$$

The results are shown in Table 6, column (3). Their inclusion has virtually no impact on any of the point estimates except for small changes for Black and parent rank. None of the changes are statistically significant. In Table 6, column (4), I include the CZ characteristics with CHKS predicted rank in place of parent rank. In this case, the point estimates for Black (interacted and uninteracted with predicted rank) increase but the differences are not statistically significant due to the large standard errors.

Next, I test whether the associations found by CHKS are present in the CSD data. To do so, I regress child rank on parent rank and a vector of CZ-level information (z_{ic}) for each individual i as

$$y_i = \beta_x x_i + \beta_c z_{ic} + \beta_{c,slope} x_i z_{ic} + e_i. \tag{4.4}$$

I have chosen to test the model at the individual level rather than at the CZ-level as in CHKS due to the small number of CZs with a sufficiently large population in the CSD sample. Each CZ-level characteristic has been normalized so that it has a mean of zero and standard deviation of 1.[23]

In Table 9, columns (1) and (2), I report the CZ-level correlations from CHKS between their two mobility measures and each of the aforementioned five main factors. Each factor at the CZ level is highly correlated with upward mobility, with a negative relationship for inequality, school quality, and single parenthood and a positive relationship for lower spatial mismatch and greater social capital. To test for the CHKS results using the individual CSD data, I regress child rank on parent rank and each of the five aforementioned characteristics separately, as an intercept term (uninteracted, to test for the correlation with absolute mobility).[24] The results are shown in column (3). For each factor except fraction with a short commute, the coefficients are statistically significant in the expected direction (at 10% for high school dropout rate). In column (4), I report the coefficients for each factor regressed separately as both an intercept and slope term (interacted with parent rank, to test for the correlation with relative mobility). The intercept term is significant for all factors but commute time and the social capital index and the slope is significant only for the rate of single parenthood. All of the signs are in the expected direction. The results from (3) and (4) generally correspond to those found by CHKS.

In column (5), I include the slope and intercept terms for all the factors simultaneously. In this regression, the rate of single parenthood is negatively correlated with upward mobility in both the slope and intercept terms (10% significance for the slope). The results in columns (4) and (5) echo the CHKS finding that single parenthood has the strongest association with low mobility across their different regression models and sample definitions.[25] In each case, the magnitudes on the coefficients are large. For example, in model (4), a one standard deviation increase in the fraction of children born to single parents is associated with a 2.14 percentile decline in expected rank of children. Using data from the CZ-groups in Figure 4 Panel B, that is approximately equivalent to moving to a city 21 percentile less mobile in the CZ distribution of the intercept. The same increase in the rate of single parenthood is also associated with a move to a city 17 percentile less mobile in the CZ distribution of the slope.

[23] As in CHKS, the normalization is done at the CZ level. I include a dummy for the presence of data for a given characteristic at the CZ level, as some indicators are missing for some CZs.
[24] The CZ-level information is taken directly from the tables made available by CHKS on their website at http:// http://www.equality-of-opportunity.org/.
[25] See Table VI and online Table VIII in CHKS.

In columns (6)-(8) I add the baseline family and demographic characteristics to the regressions in columns (3)-(5) In column (6), the intercepts considered separately are significant for fraction with a short commute only, but the sign is in the wrong direction. The point estimates on each other characteristic is reduced by nearly a third or more from (3). In column (7), I test the factors considered separately as both as intercept and slope terms simultaneously. In none of the cases are the coefficients statistically significantly different from zero at the 90% level and in some cases the sign is in the wrong direction.[26] This is in contrast to the analogue in column (4) without the baseline model characteristics where the majority of the intercept terms are significant and all of the coefficients have the expected sign.

In column (8), I test all of the variables simultaneously. In this model, none of the variables is significant with the expected sign. For only high school dropout rate are the signs of both coefficients in the expected direction. Again, this contrasts with column (5) where the fraction born to single parents is significant for the slope and intercept.[27]

The results in columns (3)-(5) confirm that many of the relationships found by CHKS between CZ-level characteristics and upward mobility are present in the CSD data. The results in columns (6)-(8) suggest that these correlations may be due to unobserved individual-level characteristics, such as race and parent education.

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 investigate heterogeneity in 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. In conditional quantile regressions, high vs. low achievement is defined by comparing children with the same characteristics. For example controlling for parent rank, education, and family type, do high-achieving Hispanic children have earnings ranks that are greater than high-achieving White children?

[26] This includes the intercept for fraction with a short commute, the GINI, and the social capital index and the slope for the GINI and the social capital index.
[27] Because of the somewhat wide standard errors, I cannot reject that nearly all of the coefficients are not different before and after the inclusion of the family and demographic characteristics, even at the 10% level.

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 in CHKS using data they have made available on their website.[28] The results are shown in Figure 6. In both data sets, there is considerable heterogeneity in child outcomes by conditional quantile with a nearly identical pattern. For children in low quantiles, or "low-achievers" conditional on parent earnings, there is much less impact of parent rank on child rank. For children in high quantiles, or "high-achievers", there is similarly a much smaller impact of parent rank on 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 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.

Next, I regress child rank on parent rank and the family and demographic characteristics of the baseline model (Table 5, column 3) with a conditional quantile regression and cohort weights, shown in Figure 7. This can help identify how each characteristic affects child rank across the conditional distribution. For example, the coefficient on Black is more negative at higher quantiles than lower ones, which means that there is a larger gap between high-achieving Black children and high-achieving White children than for low-achievers, conditioning on the other characteristics. The results for children of teen parents are even starker; it is the otherwise high-achieving children that are most negatively affected by being the child of a teen parent. Also notable is that for Hispanics the results differ from the OLS. Hispanic children across the distribution outperform White children by about 2 percentile points (from the 25th to 85th quantiles). In the baseline OLS model in Table 5, column 3, 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. It is worth noting that 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.

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

I also use unconditional quantile regressions to analyze the upward mobility of children conditional on parent rank and family characteristics. 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 ranks be affected by changing the sample distribution of the family and demographic characteristics?

For example, if conditional on the observed characteristics, high achieving Hispanics do outrank high achieving Whites, where in the marginal or *unconditional* distribution is this effect present? Since the conditional quantile regression controls for these characteristics, it is possible that the high-achieving Hispanics could be anywhere in the marginal distribution of child earnings. If the 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 parents with a college education, their relative success would be concentrated higher in the child earnings distribution.

The unconditional quantile regression results are shown in Figure 8. The adverse outcomes for Blacks affect children throughout the distribution, but are especially concentrated in the middle (from the 40^{th} to 60^{th} percentile). For Hispanic children, the greater upward mobility they experience in the conditional quantile regression primarily affects children who are in the bottom half of the marginal distribution. The effects of parent education differs by education type. Not surprisingly, children whose parents did not graduate high school are primarily affected at lower ranks in the child earnings distribution, whereas children from college educated parents are affected at higher earnings levels.

For children of single parents, there is no positive impact of single parenthood (after accounting for the lower earnings of one-parent families with the parent rank variable) on children at the bottom of the income distribution (< 30^{th} percentile). It is for otherwise higher earning children that are affected positively. Family earnings for single parents are lower than for two-parent households, in part due to the absence of a potential second earner. While children of single parents earn less than children of two-parent families, in the CSD data, the difference is smaller than the gap in parent earnings would predict. This could also be due to error in the measurement of single parent status, as it is set from the first observation of the family and left unchanged for all parent earnings years.

6 Conclusion

There is tremendous spatial variation in intergenerational upward mobility in the United States. Understanding the sources of this variation is important to understanding how to foster upward mobility and equality of opportunity. In this paper, I use data from Census surveys linked to administrative tax data to explore to what extent spatial variation in parent family characteristics and demographics can account for the observed pattern of spatial variation in mobility.

Parent education and child race are highly predictive of child earnings rank even after controlling for parent earnings. These characteristics are also highly correlated 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. However, the results in this paper suggest that the lower earnings of Black children or the higher earnings of children of more educated parents are not due to the places in which they live.

This paper also shows that an important share, but by no means all, of the spatial variation in mobility can be accounted for by the variation in child race, parent education, and parent family type (single or teen parenthood). In addition, many of the local area characteristics that are highly correlated with mobility, such as the share of single parents, income inequality, and social capital, are not predictive of mobility after controlling for these family and demographic characteristics. This leaves unanswered the question of what unobserved characteristics of the families or the places does account for the remaining spatial variation in mobility.

As an additional step to understanding the variation in child outcomes, I explore the heterogeneity in the relationship between parent characteristic and child earnings using quantile regressions. Parent earnings are most associated with child earnings rank for “average” children, as the lowest and highest achievers would generally be unsuccessful or successful regardless of their parents’ position in the distribution. In addition, for Black children and children of teen parents, the negative association with child earnings rank is especially strong for high achievers as well. Furthermore, Hispanic children outperform Whites, and the effect is concentrated entirely in children with low earnings as adults. Parent education also seems especially important to children in the middle of the distribution and higher.

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, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The 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.

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

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.

Putnam, Robert D. 1995. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy* 6 (1): 65–78.

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 (Year Child Age = 29)	**Match Rate (%)** CPS ASEC	SIPP	Total
2001		70.4%	70.4%
2002	77.4%	73.2%	75.4%
2003	77.5%	73.7%	75.5%
2004	77.9%	77.0%	77.4%
2005	72.0%	77.3%	74.4%
2006	67.4%	77.2%	73.4%
2007	71.8%	75.9%	74.1%
2008	70.1%	77.5%	73.7%
2009	68.6%	77.7%	73.6%
2010	65.0%	78.6%	71.6%
Total	70.6%	76.8%	73.9%

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) CPS ASEC (1)	SIPP (2)	Total (3)	Matched Parent-Child Pairs CPS ASEC (4)	SIPP (5)	Total (6)
Parent Family Earnings				66,330	65,257	65,765
Individual Child Earnings				33,190	35,613	34,465
Black	14.4%	14.4%	14.4%	12.8%	13.3%	13.0%
Hispanic	11.3%	10.8%	11.0%	9.0%	10.4%	9.7%
Highest Educated Parent						
< High School	11.8%	14.3%	13.3%	10.3%	12.8%	11.6%
High School	32.9%	34.1%	33.6%	32.8%	33.3%	33.1%
Some College	27.4%	27.8%	27.7%	28.3%	29.2%	28.8%
College or Graduate	27.8%	23.7%	25.4%	28.6%	24.7%	26.6%
Teen Parent	7.6%	7.8%	7.7%	7.2%	7.1%	7.2%
Single Parent	26.1%	34.2%	30.7%	25.1%	32.4%	28.9%
Partner Family	2.6%	2.4%	2.5%	2.3%	2.1%	2.2%
Observations	26,733	30,104	56,837	18,878	23,112	41,990

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	0.245	0.700
Males Only	0.313	0.290	0.830
Females Only	0.249	0.202	0.573

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 0.700 which is nearly equal to the ratio of the CPS-SIPP/DER slope to the baseline CHKS slope $\left(\frac{0.245}{0.341} = 0.718\right)$. The regressions in columns (2) and (3) use the cohort weights discussed in Section 2.1.

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

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

Min Observations in CPS-SIPP/DER For Inclusion	# of CZs	Unweighted Intercept	Unweighted Slope	CSD Observation Weighted Intercept	CSD Observation Weighted Slope
3	528	0.22	0.11	0.40	0.22
50	221	0.52	0.36	0.55	0.47
100	112	0.44	0.37	0.56	0.55
250	32	0.69	0.76	0.73	0.79
500	9	0.79	0.77	0.86	0.86
1000	3	1.00	0.99	1.00	0.99

B. Grouping CZs by Quantile of Rank-Rank Regression Slope

Quantile Groups	Intercept	Slope	Parent-Child Pairs in Smallest Quantile Group	Average Number of Parent-Child Pairs
5	0.99	0.99	5,598	8,315
	(0.01)	(0.02)		
10	0.98	0.96	2,629	4,157
	(0.02)	(0.04)		
20	0.95	0.92	861	2,079
	(0.03)	(0.05)		
25	0.94	0.87	547	1,663
	(0.03)	(0.05)		
50	0.87	0.79	162	831
	(0.04)	(0.06)		

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 41,190 observations over 570 CZs, or about 74 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 831 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 Earnings Mobility and Family Characteristics

Dependent Variable = Child Rank **Variable**	**Race and Parent Rank** **(1)**	**Parent Rank Interactions** **(2)**	**Baseline Model** **(3)**
Parent Rank	0.212***	0.222***	0.221***
	(0.010)	(0.016)	(0.014)
Black	-10.50***	-8.82***	-8.83***
	(0.95)	(0.95)	(0.95)
Hispanic	-1.84	1.73	1.72
	(1.44)	(1.57)	(1.57)
Female		-4.79***	-4.79***
		(0.80)	(0.80)
Highest Educated Parent			
< High School		-6.95***	-7.02***
		(1.13)	(1.11)
Some College		1.94**	1.78***
		(0.96)	(0.50)
College+		7.87***	7.80***
		(1.16)	(1.16)
Teen Parent		-3.06***	-3.06***
		(0.69)	(0.69)
Unmarried Partners		-5.42***	-5.42***
		(1.23)	(1.23)
Single Parent		1.91***	1.91***
		(0.61)	(0.61)
Interacted With Parent Rank			
Black	0.084***	0.062***	0.062***
	(0.022)	(0.022)	(0.022)
Hispanic	0.014	-0.007	-0.007
	(0.036)	(0.037)	(0.037)
Female		-0.083***	-0.083***
		(0.015)	(0.015)
Highest Educated Parent			
< High School		0.043	0.045
		(0.030)	(0.029)
Some College		-0.004	
		(0.020)	
College+		-0.033*	-0.031*
		(0.018)	(0.018)
Constant	41.80***	43.60***	43.68***
	(0.60)	(0.94)	(0.92)
R-Squared	0.06	0.10	0.10
Observations	41,990	41,990	41,990

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 Section 2.1. 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. Although none of the characteristics interacted with rank is significantly different from zero in (2), 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.

Table 6: Intergenerational Mobility and Family Characteristics Controlling for Spatial Heterogeneity

Dependent Variable = Child Rank	**Baseline**		**With CZ Characteristics**	
	Parent Rank	**CHKS Predicted Rank**	**Parent Rank**	**CHKS Predicted Rank**
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Rank (Parent or CHKS Predicted)	0.221***	0.635***	0.205***	0.676***
	(0.014)	(0.041)	(0.019)	(0.050)
Black	-8.83***	-9.46***	-9.24***	-12.80***
	(0.95)	(2.63)	(1.03)	(2.50)
Hispanic	1.72	-2.18	1.81	-3.44
	(1.57)	(5.14)	(1.46)	(5.01)
Female	-4.79***	3.40*	-4.94***	3.48*
	(0.80)	(1.94)	(0.79)	(1.92)
Highest Educated Parent				
< High School	-7.02***	-13.47***	-7.23***	-14.67***
	(1.11)	(3.21)	(1.11)	(3.13)
Some College	1.78***	1.74***	1.74***	1.65***
	(0.50)	(0.49)	(0.50)	(0.49)
College+	7.80***	11.96***	8.06***	12.92***
	(1.16)	(2.72)	(1.17)	(2.69)
Teen Parent	-3.06***	-3.10***	-3.05***	-2.99***
	(0.69)	(0.68)	(0.69)	(0.68)
Unmarried Partners	-5.42***	-5.53***	-5.59***	-5.47***
	(1.23)	(1.24)	(1.24)	(1.22)
Single Parent	1.91***	1.85***	1.64***	1.87***
	(0.61)	(0.59)	(0.62)	(0.60)
Interacted With Rank (Parent or CHKS Predicted)				
Black*Rank	0.062***	0.09	0.066***	0.14**
	(0.022)	(0.06)	(0.022)	(0.05)
Hispanic*Rank	-0.007	0.073	-0.002	0.096
	(0.037)	(0.111)	(0.036)	(0.111)
Female*Rank	-0.083***	-0.248***	-0.082***	-0.250***
	(0.015)	(0.041)	(0.015)	(0.040)
< High School*Rank	0.045	0.185**	0.049*	0.215***
	(0.029)	(0.074)	(0.029)	(0.072)
College+*Rank	-0.031*	-0.108**	-0.036**	-0.134***
	(0.018)	(0.050)	(0.018)	(0.050)
Constant	43.68***	22.70***	43.34***	18.28***
	(0.92)	(2.15)	(0.97)	(2.60)
CHKS CZ Characteristics			X	X
R-Squared	0.10	0.11	0.11	0.11
Observations	41,990	41,573	41630	41,573

In this table, I test the weighted OLS regressions of child rank on parent rank (models (1) and (2)) and on predicted child rank (models (3) and (4)). 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 Section 2.1 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 Putnam (Putnam 1995)), and the fraction of single mothers. ***, **, * represent significance at the 0.01%, 0.05% and 0.10% level.

Table 7: Correlation Between Family and Demographic Characteristics and CHKS Mobility

A. By CZ Groups

CZ Groups	5		10		20		25		50	
CHKS Coefficient	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Black	-0.91	0.88	-0.90	0.87	-0.86	0.87	-0.82	0.84	-0.47	0.40
Hispanic	0.85	-0.89	0.70	-0.77	0.63	-0.70	0.32	-0.48	-0.09	-0.24
Parent Education										
< High School	0.17	-0.25	0.07	-0.19	-0.14	0.07	-0.38	0.26	-0.22	-0.10
High School	-0.89	0.92	-0.75	0.83	-0.66	0.77	-0.66	0.76	-0.46	0.66
Some College	0.94	-0.94	0.68	-0.68	0.58	-0.64	0.66	-0.67	0.44	-0.27
College +	0.81	-0.73	0.56	-0.51	0.58	-0.56	0.64	-0.61	0.29	-0.24
Teen Parent	-0.90	0.84	-0.85	0.81	-0.83	0.81	-0.82	0.80	-0.54	0.41
Single Parent	-0.73	0.68	-0.67	0.64	-0.72	0.65	-0.75	0.63	-0.69	0.45
Unmarried Partner	0.50	-0.56	0.23	-0.25	-0.13	0.02	-0.02	-0.07	0.00	-0.05

Low Mobility **High Mobility**

B. Regression of CHKS Mobility Measures on Family and Demographic Characteristics

Mobility Measure	Intercept		Slope	
	Individually	Together	Individually	Together
	(1)	(2)	(3)	(4)
Black	-3.28***	-3.05***	3.59***	3.29***
	(0.38)	(0.38)	(0.43)	(0.40)
Hispanic	1.92***	2.07***	-4.50***	-4.62***
	(0.49)	(0.44)	(1.09)	(1.02)
Parent Education				
< High School	-0.45	-0.41**	-0.45	-0.04
	(0.39)	(0.19)	(0.70)	(0.26)
High School	-0.45***		1.04***	
	(0.14)		(0.25)	
Some College	0.39***	0.53***	-0.57***	-1.10***
	(0.13)	(0.15)	(0.15)	(0.22)
College +	0.34**	0.31*	-0.35*	-0.86***
	(0.13)	(0.17)	(0.20)	(0.25)
Teen Parent	-1.11***	-0.27**	1.09***	0.37**
	(0.16)	(0.13)	(0.20)	(0.16)
Single Parent	-0.73***	-0.12	0.46***	-0.16
	(0.14)	(0.10)	(0.14)	(0.13)
Unmarried Partner	0.17	0.26	-0.14	-0.31
	(0.23)	(0.22)	(0.27)	(0.26)
Constant		33.48***		34.17***
		(0.32)		(0.42)
Observations	41,573	41,573	41,573	41,573

Panel A shows the correlation between the baseline model family and demographic characteristics and CHKS CZ-level mobility. 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. The correlations shown are unweighted correlations with the observation-weighted CHKS intercept α_c^{CHKS} and slope β_c^{CHKS} from the CZ-level mobility regression in equation (3.3). 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. For the intercept, -1 is perfectly correlated with low mobility and 1 is perfectly correlated with high mobility with the reverse for the slope. Panel B columns (1) and (3) show the CHKS intercept and slope regressed against each characteristic separately at the individual level. Columns (2) and (4) show the measures regressed against all the characteristics simultaneously. In Panel B, each regression uses the cohort weights discussed in Section 2.1, and the errors are clustered at the CZ level. ***, **, * represent significance at the 0.01%, 0.05% and 0.10% level.

Table 8: Variation Explained by Family and Demographic Characteristics

Measure	CZ Quantile Groups 5	10	20	25	50
Regression Coefficient					
Intercept ($\alpha_{p,adjusted} = \gamma_\alpha + \delta_\alpha \alpha_p$)	0.48	0.47	0.50	0.50	0.58
Slope ($\beta_{p,adjusted} = \gamma_\beta + \delta_\beta \beta_p$)	0.63	0.64	0.66	0.68	0.72
Measures of Dispersion					
Variance					
Intercept	0.70	0.71	0.66	0.63	0.54
Slope	0.55	0.52	0.48	0.44	0.38
Coefficient of Variation					
Intercept	0.50	0.51	0.47	0.45	0.38
Slope	0.28	0.26	0.22	0.20	0.16
Mean Absolute Deviation					
Intercept	0.47	0.47	0.40	0.41	0.35
Slope	0.36	0.31	0.27	0.23	0.22

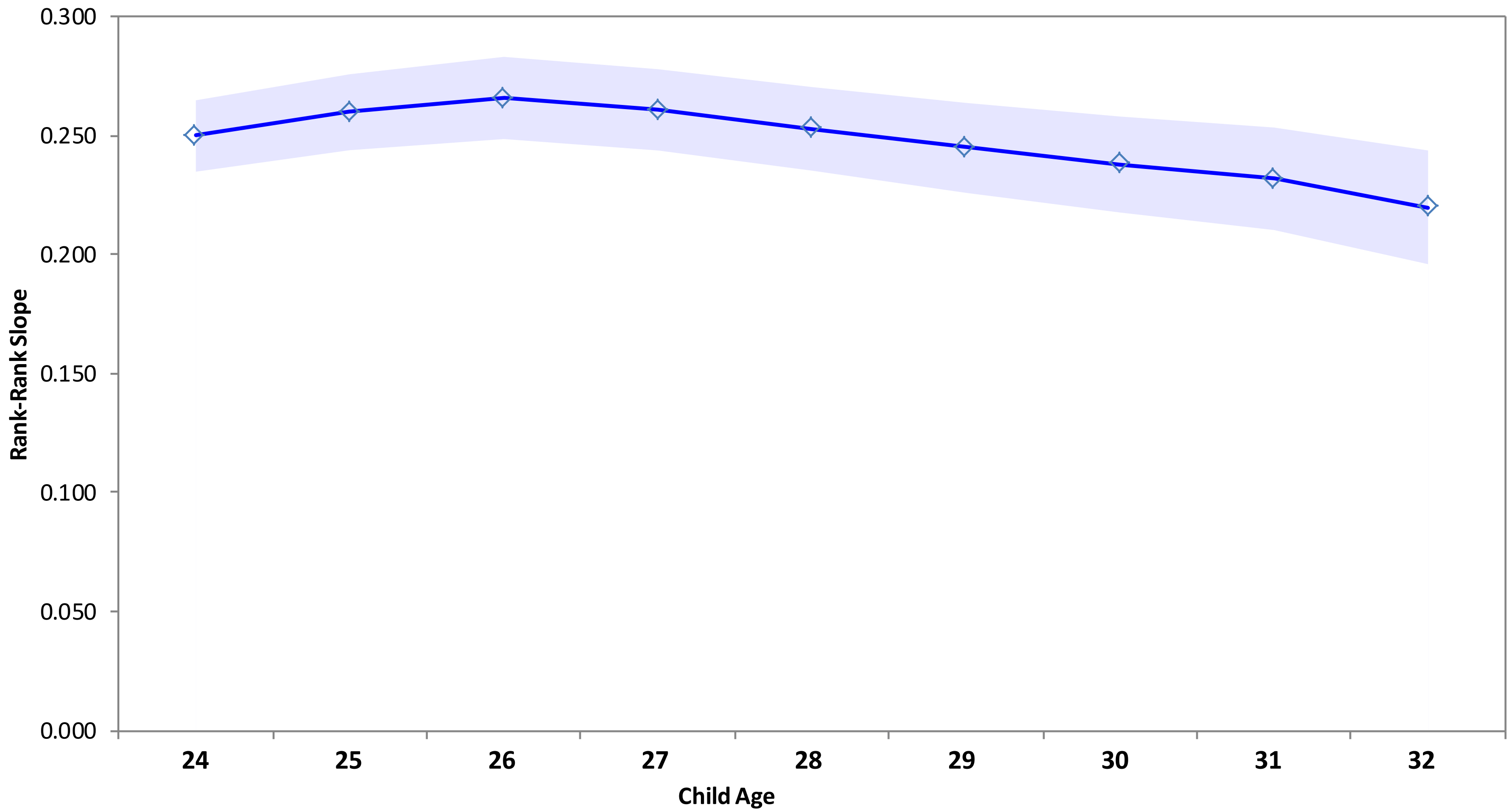
In this table, I test the variation explained by the demographic and family characteristics by regressing child rank against the baseline model (model (3) in Table 5) and creating an adjusted child rank where $y_{i,Adjusted} = y_i - \beta_h h_i - \beta_{xh} x_i h_i$ (using cohort weights). The $y_{i,adjusted}$ accounts for the relationship between the observable characteristics and child rank but not parent rank. As before, I divide the individuals into k ($p = 1, \ldots, k$) CZ quantile groups based on the relative mobility in each CZ. I then calculate the CZ-group level slope and intercept from the regression of unadjusted child rank $y_{ic} = \alpha_p + \beta_p x_{ic}$ and adjusted child rank $y_{ip,adjusted} = \alpha_{p,adjusted} + \beta_{p,adjusted} x_{ip} + e_{ip,adjusted}$. I then regress the adjusted slope and intercept coefficients on the unadjusted ($\alpha_{p,adjusted} = \gamma_\alpha + \delta_\alpha \alpha_p$ and $\beta_{p,adjusted} = \gamma_\beta + \delta_\beta \beta_p$) to determine to what extent the two differ as a result of the inclusion of the individual and family demographic characteristics. 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 ($1 - \frac{\text{unadjusted}}{\text{adjusted}}$).

Table 9: Commuting Zone Characteristics and Intergenerational Mobility

	CHKS		CPS-SIPP/DER Sample										
			Without Family Variables					With Baseline Family Variables					
			Each CZ Characteristic Separately			CZ Characteristics Together		Each CZ Characteristic Separately			CZ Characteristics Together		
	(1)	(2)	(3)	(4)		(5)		(6)	(7)		(8)		
CZ Characteristic	Correlation with AM	Correlation with RM	Intercept Only	Intercept	Slope	Intercept	Slope	Intercept Only	Intercept	Slope	Intercept	Slope	
Fraction With Commute < 15 Min	0.61***	-0.45***	-0.71*	-1.14*	-0.003	-1.04	-0.011	-1.28***	-1.14*	-0.003	-1.34**	-0.010	
	(0.13)	(0.07)	(0.40)	(0.59)	(0.013)	(0.71)	(0.015)	(0.45)	(0.59)	(0.013)	(0.64)	(0.015)	
GINI of Bottom 99%	-0.65***	0.47***	-0.95**	-1.48***	0.011	-0.82	-0.001	-0.31	0.01	-0.006	-0.29	-0.008	
	(0.09)	(0.09)	(0.37)	(0.56)	(0.010)	(0.88)	(0.016)	(0.38)	(0.47)	(0.009)	(0.78)	(0.014)	
High School Dropout Rate	-0.57***	0.33***	-0.75*	-1.85***	0.024	-1.22	0.015	-0.25	-0.66	0.009	-0.90	0.010	
	(0.09)	(0.10)	(0.41)	(0.67)	(0.015)	(0.89)	(0.018)	(0.41)	(0.63)	(0.015)	(0.71)	(0.015)	
Social Capital Index	0.64***	-0.33***	0.90**	1.02	-0.003	-0.13	0.015	0.32	-0.20	0.011	-0.16	0.014	
	(0.09)	(0.09)	(0.35)	(0.68)	(0.014)	(0.84)	(0.017)	(0.37)	(0.56)	(0.012)	(0.80)	(0.017)	
Fraction Born to Single Parents	-0.76***	0.64***	-0.90*	-2.14***	0.027***	-1.57**	0.029*	0.09	0.00	0.002	0.12	0.009	
	(0.07)	(0.05)	(0.49)	(0.56)	(0.010)	(0.77)	(0.015)	(0.46)	(0.55)	(0.011)	(0.70)	(0.015)	

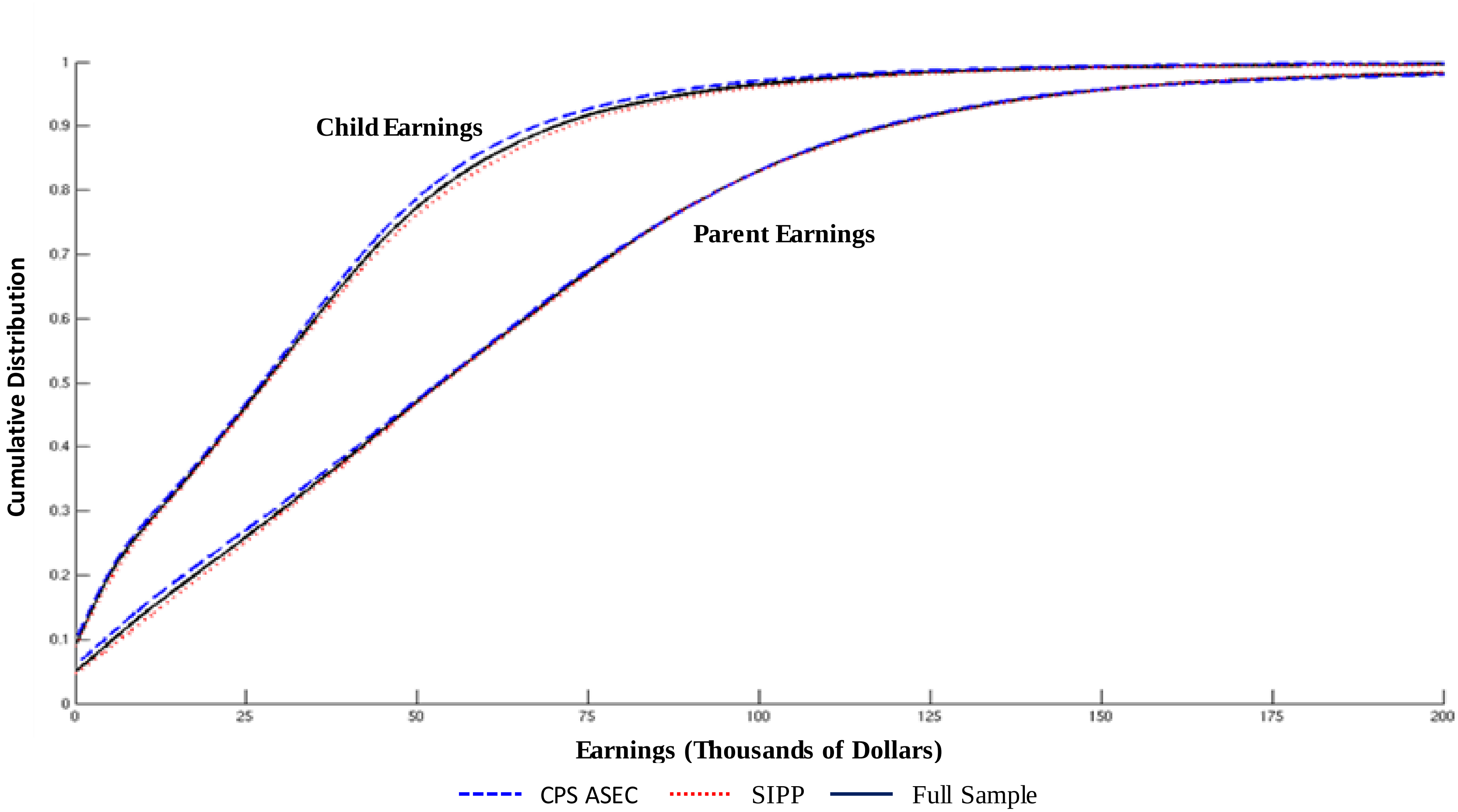
Columns (1) and (2) show the correlation between the absolute mobility (AM) and relative mobility (RM) measures and CZ level characteristics reported in CHKS. In Columns (3)-(6), each CZ characteristic was included as an independent variable in a separate regression of child rank on parent rank. In column (3) only the CZ characteristic was included and in (4), the CZ characteristic and its interaction was included in the regression. In (5), all the characteristics were included simultaneously in a regression of child rank on parent rank. Columns (6)-(8) report the same coefficients as (3)-(5), but in a regression that also included the baseline family variables of race, parent education, and single and teen parenthood. All regressions were run using the cohort weights discussed in Section 2.1 with errors 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



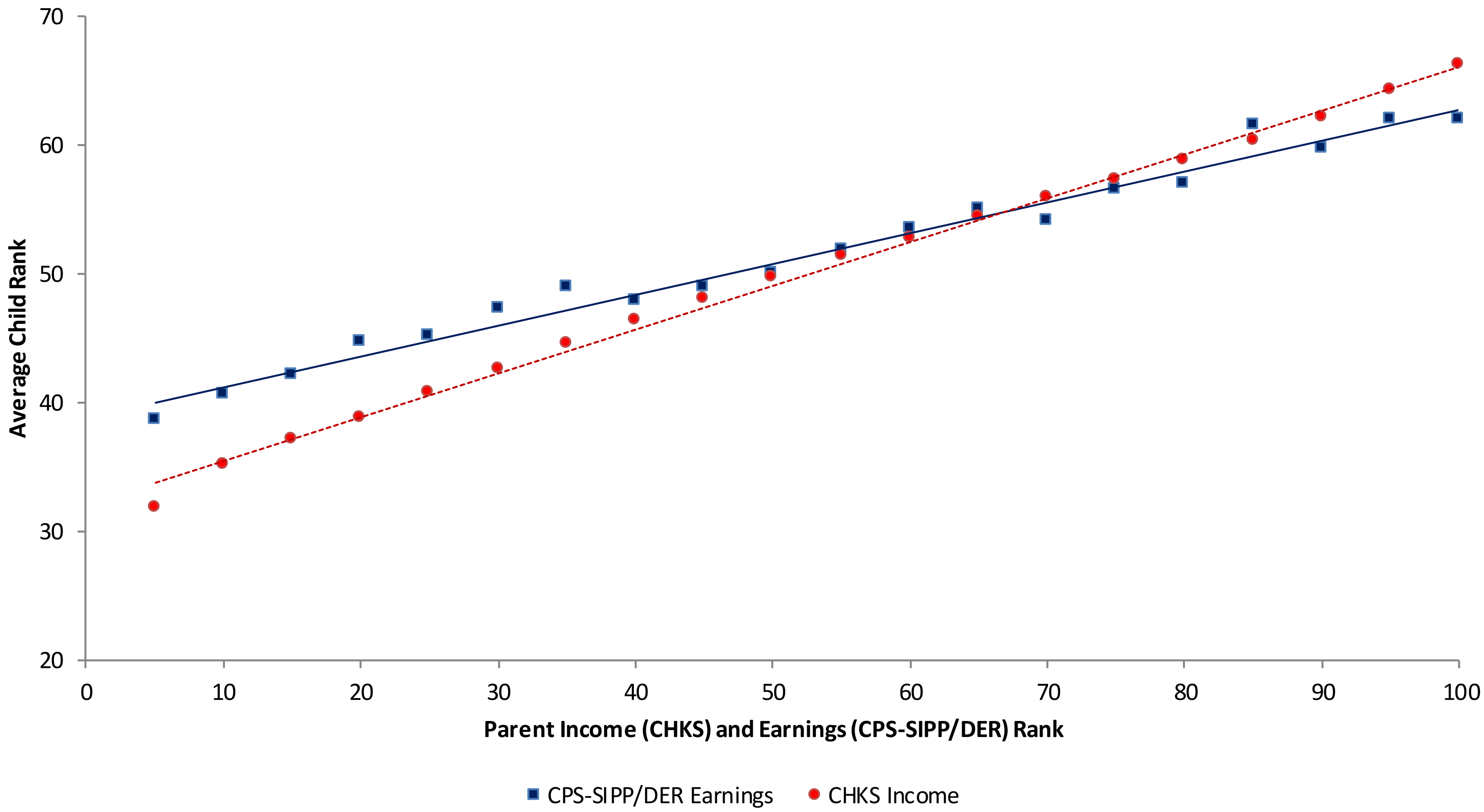
This figure plots the rank-rank slope of intergenerational mobility. The parent family earnings are the average when the older parent is 40-44 years old. The child earnings are the average of earnings for age t and $t+1$ where t varies from 24 to 32. It should be noted that as t varies, so does the size 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 20%. For example, at $t=32$, there are 21,855 parent-child pairs and at $t=24$, there are 92,188 parent-child pairs.

Figure 2: 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.

Figure 3: Association between Parent and Child Rank in CHKS and CPS-SIPP/DER

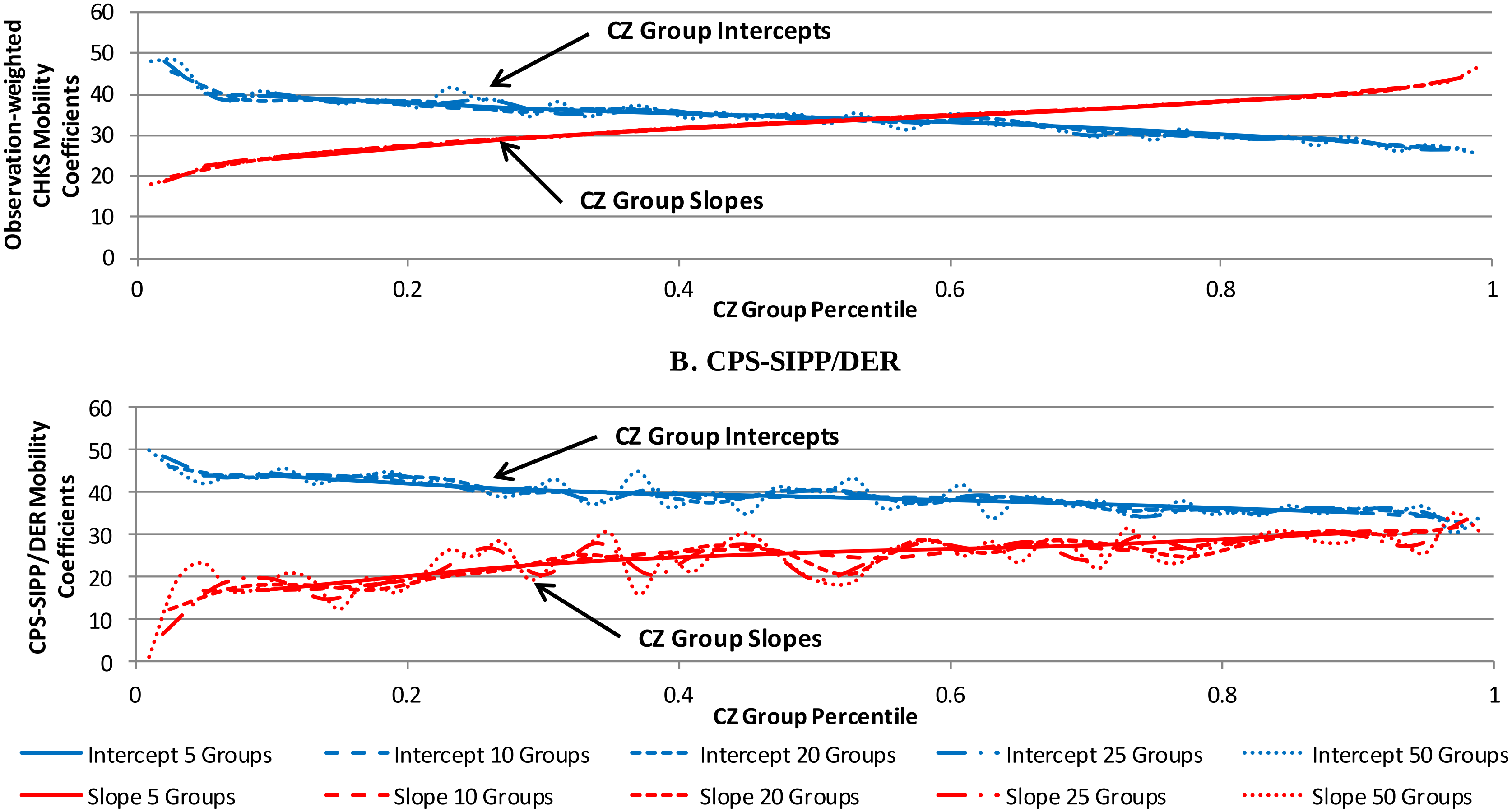


This figure plots the average child rank for 20 parent rank bins. In both the CPS-SIPP/DER and the CHKS data, the relationship between parent rank and average child rank is very well represented by the linear regression slope and intercepts on the individual observations. The CHKS parent ranks are determined 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 4: Mobility Parameters for Each CZ Quantile Group

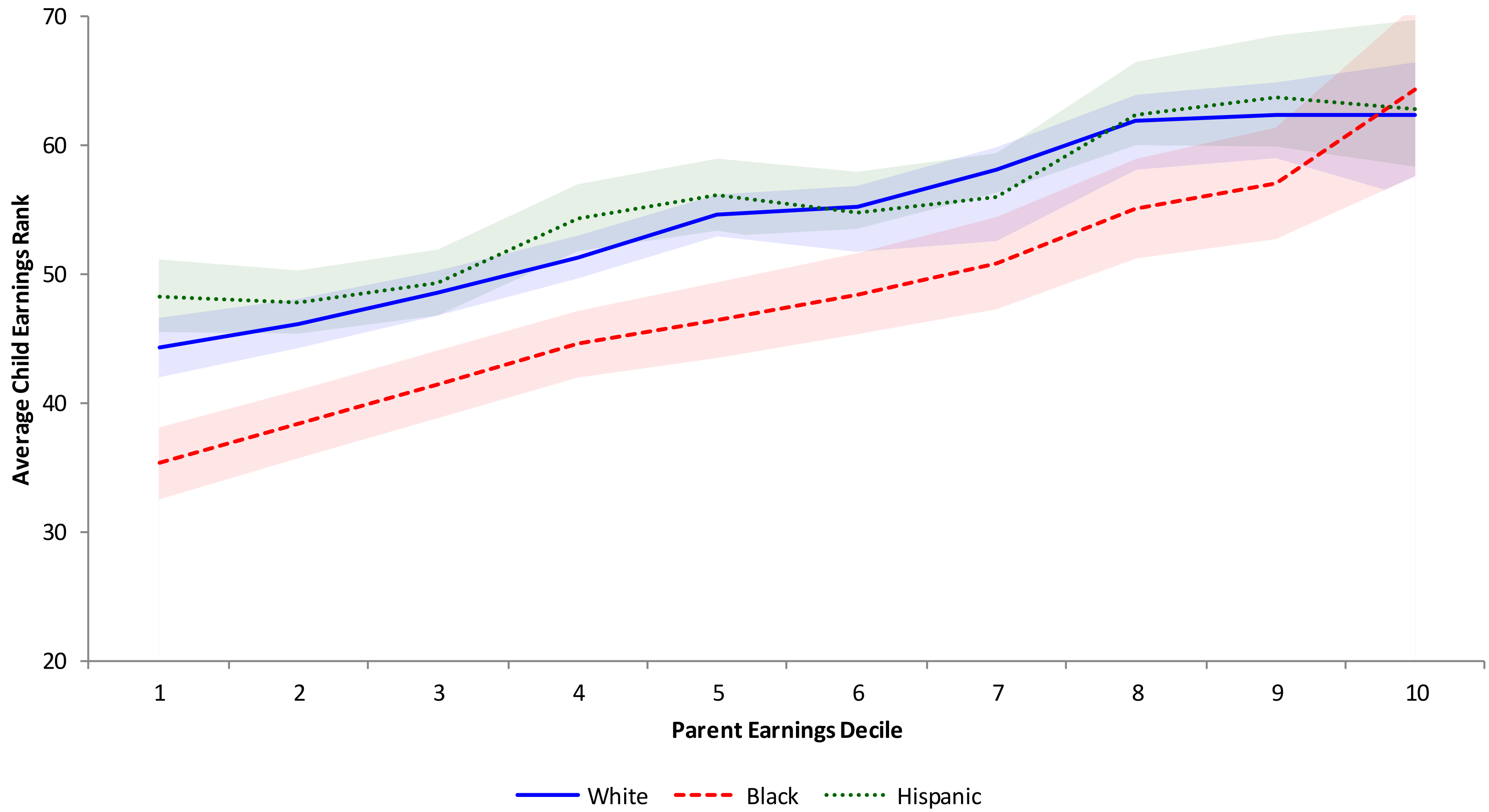
A. CHKS

B. CPS-SIPP/DER



This figure plots the average CHKS and CPS-SIPP/DER CZ quantile group mobility coefficients, the regression intercept (α_p) and slope (β_p). The lines are plotted with the CZs divided into 5, 10, 20, 25, and 50 equally sized groups based on the CZ slope from most to least mobile. In Panel A, the quantile group coefficients are the CPS-SIPP/DER observation-weighted averages of the CHKS intercept and slope. In Panel B, the coefficients are from the CPS-SIPP/DER quantile group rank-rank regressions. The quantile groups are used to test for spatial heterogeneity of mobility in the CPS-SIPP/DER sample.

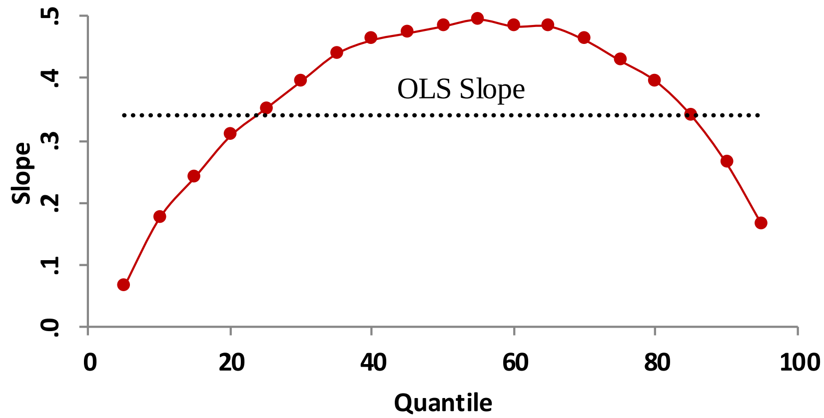
Figure 5: Race and Mobility by Decile Controlling for Parent Education



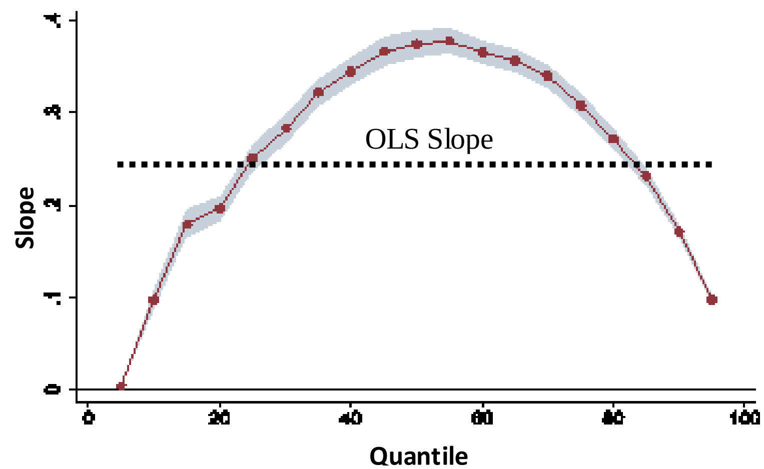
This figure plots the results of an OLS regression with dummies for each parent earnings decile interacted with race and highest parent education level (less than high school, some college, and college and above with high school as the default category). The three categories plotted by decile with 95% confidence intervals are White (and other), Black, and Hispanic.

Figure 6: Conditional Quantile Regression of Child Rank on Parent Rank

A. CHKS

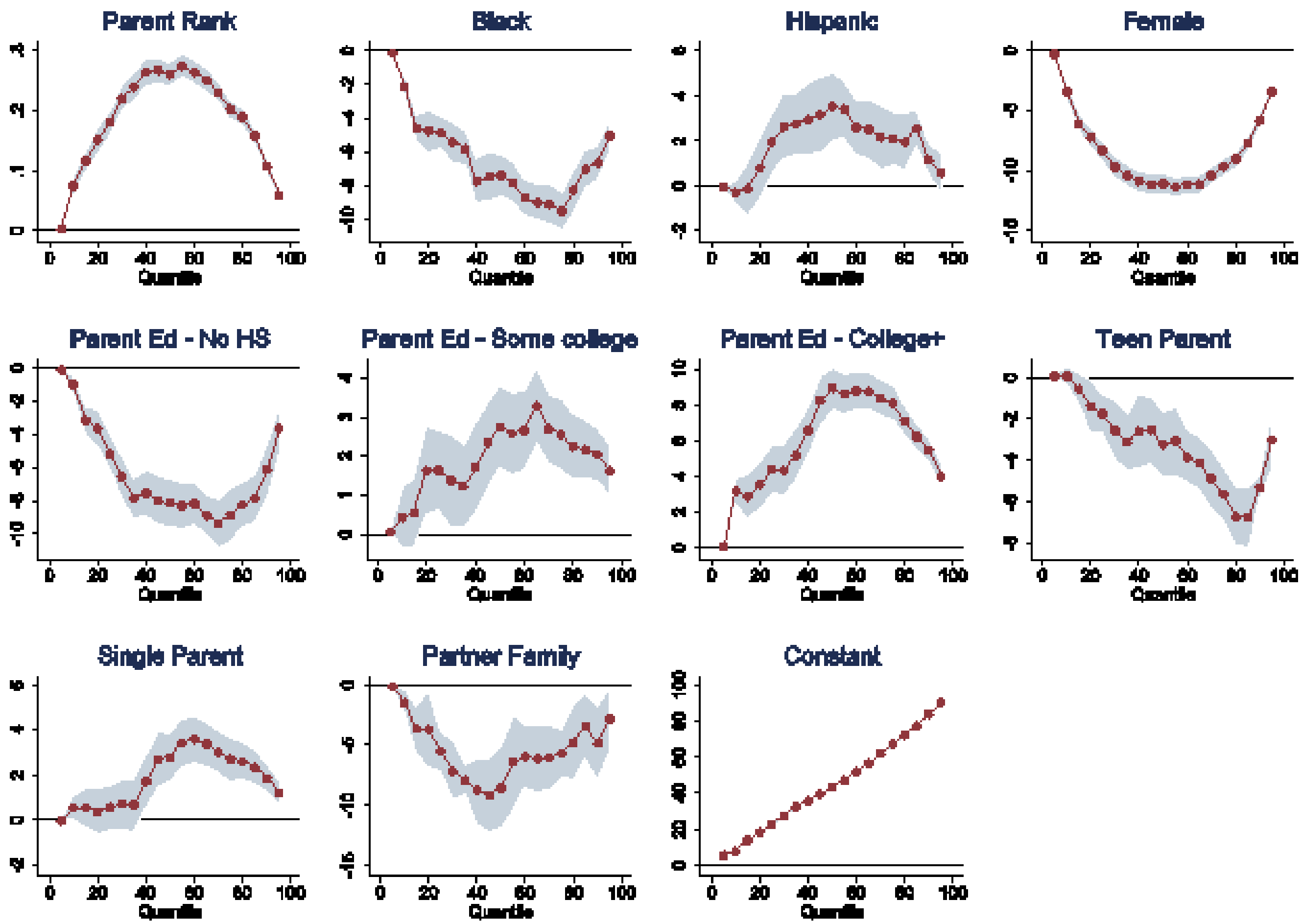


B. CPS-SIPP/DER



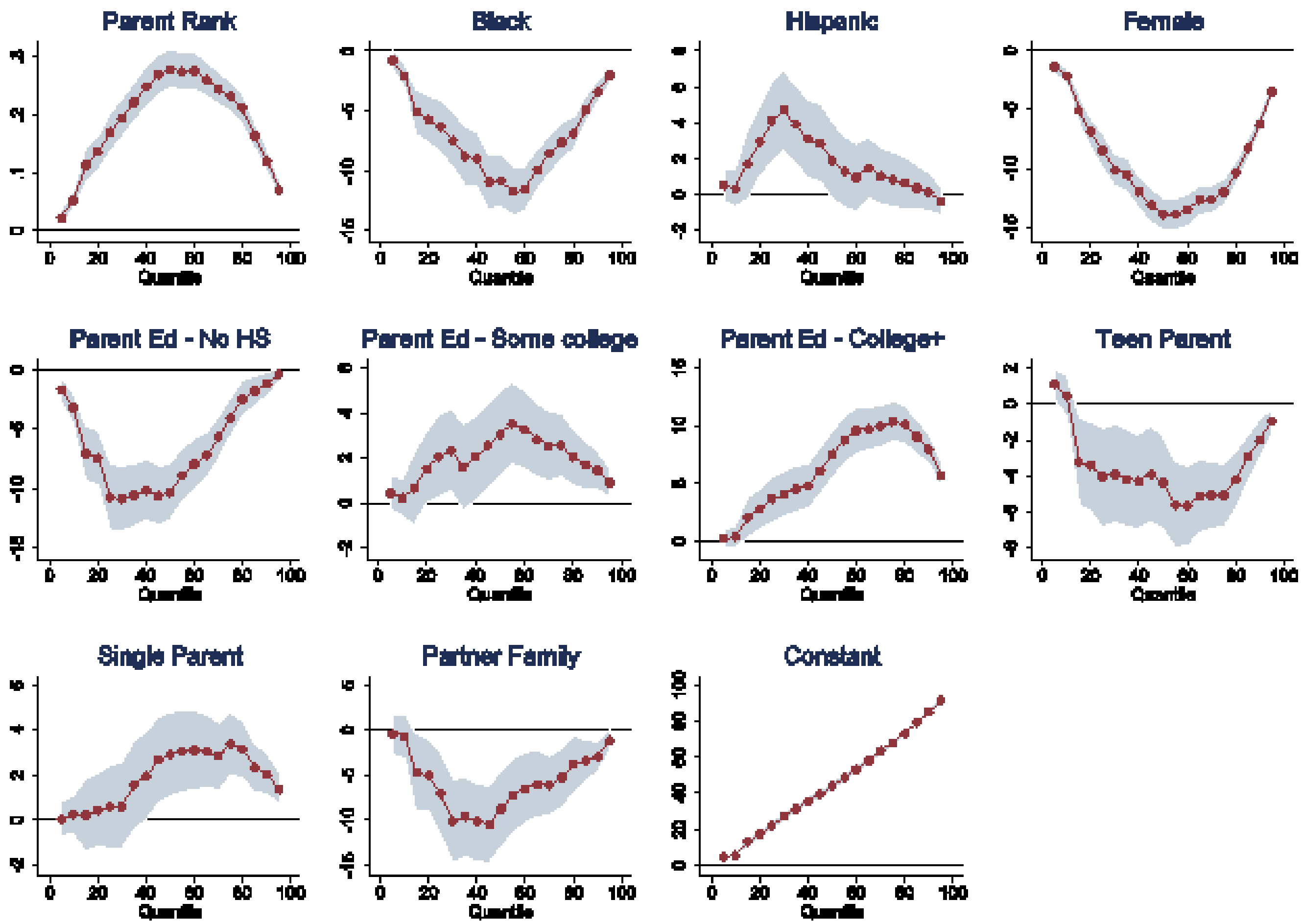
This figure shows the quantile regression of child rank on parent rank. Panel A shows the CHKS results (author's calculation from transition matrix data available at http://www.equality-of-opportunity.org/), and Panel B shows the results using the cohort-weighted CPS-SIPP/DER data with 95% confidence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

Figure 7: Conditional 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 8: Unconditional 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 upward mobility of different subgroups has an effect. The 95% confidence intervals are shown.