The College Of WILLIAM & MARY

The Relationship between Location Choice and Earnings Inequality*

Peter McHenry
College of William and Mary

College of William and Mary
Department of Economics
Working Paper Number 112

February 2011

* I thank Melissa McInerney and Florian Ploeckl for helpful comments. I am grateful for the support of a Faculty Summer Research Grant from the College of William and Mary, which partially supported this research.

COLLEGE OF WILLIAM AND MARY
DEPARTMENT OF ECONOMICS
WORKING PAPER #112
February 2011

The Relationship between Location Choice and Earnings Inequality

Abstract

This paper provides new empirical evidence about how workers' locations affect measurements of earnings inequality (and their changes over time) in the United States. Part of the inequality observed in any given U.S. sample is due to the fact that workers with different skills (and therefore earnings) are not distributed symmetrically across locations that are more and less productive (and therefore pay higher and lower wages). In particular, I estimate that a significant and rising proportion of the college wage premium is due to college graduates living in and moving toward higher-paying locations than high school graduates. Furthermore, I assess the impact of location on real wage inequality (adjusting for local costs of living). The higher wages that college graduates enjoy as a result of their location choices are mostly counterbalanced by higher costs of living. From this, I infer that college graduates choose to live in more economically productive labor markets than do workers with less education, but college graduates are not necessarily more capable of exploiting locational wage differences for their own advantage.

JEL Codes: J31, R23, J61
Keywords: Earnings inequality, Migration, Regional labor markets

Peter McHenry
Department of Economics
College of William and Mary
Williamsburg, VA 23187-8795
pmchenry@wm.edu
http://wmpeople.wm.edu/site/page/pmchenry.

1 Introduction

Similar workers in different locations often earn different wages. So, part of measured earnings inequality is due to location-specific wage determinants, like the presence of a natural port or agglomeration economies. Higher wages in a particular location may improve the welfare of its workers if local land prices do not fully compensate (say, if workers have limited mobility). A major focus of economic research has been to understand the causes and consequences of a rapid rise in U.S. earnings inequality from the 1970s to the present. Relatively few authors explicitly study how location affects earnings inequality (those few include Moretti (2010) and Black, et al. (2009)).

However, location-specific wages may have a variety of implications for inequality research. If workers with higher education are better at choosing productive locations than less-educated workers, then encouraging better mobility decisions may reduce earnings inequality. On the other hand, higher-educated workers may be more concentrated in high-earnings productive places, but higher costs of living fully compensate for their location-induced higher earnings. Then, earnings inequality measured without accounting for cost of living (the most common method) overstates inequality in people's welfare or consumption possibilities.

This paper provides new empirical evidence about how workers' locations affect measurements of earnings inequality (and their changes over time) in the United States. I analyze migration behavior in two longitudinal data sets housed at the U.S. Department of Education: the National Longitudinal Study of the High School Class of 1972 (NLS-72) and the National Education Longitudinal Study of 1988 (NELS:88). Both data sets include random samples of seniors in U.S. high schools and follow the respondents with questionnaires into their mid-20s. The NLS-72 cohort members are seniors in 1972, and the NELS:88 cohort are seniors in 1992. These data sets imply nationwide trends in early adulthood behavior from the 1970s through the 1990s, ideal for a study of determinants of the rise in U.S. earnings inequality. In particular, I investigate differences over time in migration behavior (location choice) of college graduates (the highly-educated group) and high school graduates (who did not attend college).

As the measurement of earnings inequality, I focus on the log earnings gap between college graduates and high school graduates (the "college wage premium"). I find that a significant and rising proportion of the college wage premium is due to college graduates living in and moving toward higher-paying locations than high school graduates. Furthermore, I find differences in the local costs of living where college and high school graduates choose to live: college graduates choose more expensive locations. The differences between college and high school graduates in location wages and location-specific price indices are approximately equal. From this, I infer that college graduates choose to live in more economically productive labor markets (higher wages and higher costs of living) than do workers with less education.

An implication is that nominal earnings inequality measures (not accounting for local costs of living) have over-stated the degree of consumption or welfare inequality in the U.S. (which depend on costs of living). This is the theme of Moretti (2010). He uses 1980 and 2000 U.S. Census data to calculate college wage premia that account for MSA-specific costs of living and that do not. He estimates that about 25 percent of the raw (nominal) college wage premium is accounted for by local costs of living.

Moretti (2010) also provides evidence that differential location of college graduates is related to relative demand for college-graduate labor in relatively expensive cities, rather than the alternative explanation of labor supply shocks. In particular, the locations with relatively large college premia are the locations that experienced the largest levels and increases in their populations' college shares. When prices and quantities are both increasing, demand is likely to be shifting to the right.

The findings reported here complement Moretti's. I use longitudinal data in order to distinguish between origin location (where a respondent's high school is) and destination location chosen by the respondent. This allows me to observe directly the extent to which workers use migration choices to increase their earnings. Moreover, earnings in a worker's origin are more associated with his parents' choices than his own, so this part of earnings inequality is associated with intergenerational mobility. On the other hand, earnings associated with a worker's adult location are more closely linked to his own investment choices. I find that both location in high school and as an adult contribute to overall earnings inequality, but the effect of high school location is larger. The importance of high school location indicates intergenerational persistence, a finding that is related to a large research field on neighborhood effects and parent-child earnings correlations (see Solon 1999).

My findings imply that college graduates on average do not parlay their higher education into better migration decisions in terms of finding locations where higher real earnings (and more consumption) are available, relative to workers with less education. However, it is possible that college graduates still gain in other ways from their relative concentration in more productive locations. For example, they may be more likely to be promoted throughout their careers, or they may be less prone to the negative consequences for employment of living in a declining area.

2 Economic framework: Workers' locations and the college premium

Consider an economic model with multiple local economies that each contain firms and workers producing both goods that are traded across economies (locations) and goods that are not traded. The framework follows Roback (1982) and subsequent spatial equilibrium models. Workers are mobile and move toward locations with relatively high wages and relatively low local prices (costs of living). In equilibrium, the vectors of location-specific wages and costs of living are such that workers do not want to change locations. If workers are perfectly mobile, then those with identical productive characteristics (e.g., education) attain the same real wages (wages net of cost of living) in all locations. If workers are imperfectly mobile (as in the model of Moretti (2010)), then only the marginal workers will be indifferent between locations, and average real wages may differ across locations.

Suppose locations vary in their characteristics, with some being more amenable to workers and some being more productive for firms. Amenable locations tend to have high costs of living and low wages, whereas productive locations tend to have high costs of living and high wages. Observationally equivalent workers in different locations may obtain different wages: higher wages would be offered in the more productive location if either 1) they were off-set by an equivalent increase in the cost of living (else workers would migrate there and drive wages down), or 2) workers are not perfectly mobile. The same reasons may yield higher wages in a less amenable location that is nevertheless productive enough to justify the presence of firms willing to compensate workers for living with fewer amenities.[1]

As described above, workers' different locations account for part of the variation in their wages. Following this idea, I specify a decomposition of expected wages into parts due to individual characteristics and parts associated with one's location. For simplicity, I focus on education as the only individual characteristic that affects wages.[2] The idea is that people with higher education tend to have higher earnings through higher productivity in all locations, but wages also vary across locations due to local characteristics that affect labor productivity (demand) and amenability (supply). Suppose expected log wages are:

$$E(w|S,J) = E(S\Theta_s + J\Theta_j|S,J) = E(S\Theta_s|J) + E(J\Theta_j|S), \tag{1}$$

where S is a vector of indicators for skill level, J is a vector of indicators for work location, and Θ_s and Θ_j are their associated regression coefficients.

In order to understand how location choice affects the earnings distribution, it is informative to compare childhood and adult locations (as in McHenry 2010). Workers did not

[1] If physical capital is fully mobile, more and less amenable locations may feature the same wage levels. This is a result of capital augmenting the relatively high supplies of labor in amenable locations.

[2] Empirical specifications below control for more individual traits, however.

actively choose their childhood locations, but they did–for the most part[3]–choose adult locations, at least partially based on labor market opportunities. It is much different for highly-educated people to concentrate in high-earnings locations because of their parents' choices than because of their own choices. The effect of adult location on wages can be broken down into two parts: the wage effect from where a worker grew up and the difference in wages between their adult location and their origin (the migration effect). The following equation undertakes such a decomposition:

$$E(w|S,B) = E(S\Theta_s + B\Theta_b + \Delta_m|S,B) = E(S\Theta_s|B) + E(B\Theta_b|S) + E(\Delta_m|S,B), \quad (2)$$

where S is a vector of indicators for skill level, B is a vector of indicators for birthplace, and Θ_s and Θ_b are their associated regression coefficients. Δ_m is another variable that measures the effect of location on wages; in particular, it is the difference between the locational wage effects of the chosen destination and the birthplace: the location-specific effect on wages of migration away from one's birthplace. Think of $S\Theta_s$ as the part of the wage that education determines, valued at all locations equally.[4] The effect of one's location on the wage is $B\Theta_b + \Delta_m$. The connection with Equation (1) above is that $J\Theta_j = B\Theta_b + \Delta_m$. For people who stay in their birthplace, $\Delta_m = 0$, so their wage effect comes from their birthplace alone. For people who migrate, Δ_m adds to or subtracts from $B\Theta_b$ so that what is measured is the chosen residence location's effect on wages.

In the next empirical part of the paper, I describe how much U.S. earnings inequality can be attributed to differences in childhood residence ($B\Theta_b$) and how much can be attributed to differential migration behavior, conditional on childhood residence (Δ_m).

[3]The large literature on spousal ties in migration decisions following Mincer (1978) provides one reason to suspect migration decisions are not always individually earnings-maximizing. Nevertheless, I expect adult location decisions to be much closer linked to individual earnings opportunities and preferences than childhood locations are.

[4]See Black, et al. (2009) for an equilibrium model noting that this kind of locational-invariance of the return to education is justified in a setting with homothetic preferences over local goods (housing) and other goods that are traded nationally.

The degree of earnings inequality explained by childhood location is another piece of evidence about intergenerational earnings mobility (see Solon 1999). In addition, if some workers attain much higher wages by exploiting migration opportunities better than others, then a promising strategy to attenuate earnings inequality is to encourage strategic migration behavior among those with relatively low earnings where they live. In addition to wages, workers that tend to cluster in highly-productive local economies may enjoy higher employment rates and wage growth through more productive job matching.

3 Location and the college premium

3.1 Data description

I use several data sets to analyze the relationship between location choice and the college wage premium. To document the dramatic increase of the college premium in the U.S., I use data from the 1980 and 2000 U.S. decennial Censuses (using IPUMS: Ruggles, et. al. (2010)). I also use these data to construct location-specific characteristics that describe labor productivity and costs of living. To study changes over time in early labor market geographic mobility, I use two longitudinal data sets: the National Longitudinal Study of the High School Class of 1972 and the National Education Longitudinal Study of 1988.

An important research decision concerns which location definition to use. In this study, I use the commuting zone (CZ). Each CZ approximates a local labor market, which I consider to be the smallest geographic space where most residents work and most workers reside. Tolbert and Sizer (1996) describe the identification of CZs using journey-to-work data from the 1990 Census. Each CZ is a collection of counties (or a single county) with strong commuting links among them. The CZ definition has the added feature of encompassing both rural and urban areas.[5] There are 741 CZs in the U.S. 604 of them are entirely contained by a single state, 129 of them by two states, and 8 of them by three

[5]This is the same location definition used in Autor and Dorn (2008) to study the interactions of different types of workers within labor markets.

states (e.g., Washington, D.C.). CZ populations in 2000 range from 1,193 (Murdo, SD) to 16,393,360 (Los Angeles, CA). 258 CZs contain a metropolitan statistical area.

People in this paper's framework choose destinations that have particular characteristics: wages and housing prices. I use U.S. Census data (the 1980 and 2000 5 percent samples) to calculate CZ-specific measures of these characteristics. The smallest identifiable area in the Census is the "county group" in 1980 and the public use microdata area (PUMA) in 2000. Both are Census-defined places with population no less than 100,000. This definition does not allow perfect matching of boundaries for all CZs. The method used to convert PUMA averages to CZ averages involves assigning PUMA characteristics to a CZ based on the population weight of the PUMA in the CZ. I use the same method for 1980 county groups and 2000 PUMAs.

To investigate location choices over time, I use the National Longitudinal Survey of the High School Class of 1972 (NLS-72) and the National Education Longitudinal Study of 1988 (NELS:88). From both, I take a representative sample of high school seniors in the United States: in 1972 from NLS-72 and 1992 from NELS:88.[6] Both surveys were administered by the U.S. Department of Education.

Both surveys allow identification of respondent residence locations (though access is restricted by the Department of Education). I focus in particular in both surveys on the respondent's commuting zone (CZ) as a high school senior and again at age 26 (the age of final follow-up in NELS:88). The NLS-72 data file includes county of the respondent's high school and zip code of residence at the fourth follow-up survey (fielded from October 1979 to May 1980). For the NELS:88 sample, I infer zip code of a student's school from provided data on detailed Census characteristics of the school's zip code, and the NELS:88 restricted data file includes the respondent's residence zip code for the fourth (final) follow-up survey (fielded in early 2000). I assign each county or zip code location

[6]NELS:88 started with a random sample of 8th graders, but it is possible to drop respondents who did not become high school seniors in 1992 and use an appropriate weighting scheme to analyze a representative sample of high school seniors (Curtin et al. 2002). I do so in the interest of comparability between the samples.

to its associated CZ.[7]

Other information I use from the longitudinal data sets includes sex, race, ethnicity, highest level of completed schooling, work history, and earnings. The earnings variable is the sum of all labor earnings in a year and is only counted for full-time workers who have completed their formal schooling.[8] Throughout, I use sample weights that make the NLS-72 and NELS:88 longitudinal samples representative of 12th grade students in U.S. schools in 1972 and 1992, respectively.

3.2 Variation of wages across locations

Before describing how location choices interact with earnings inequality, I provide baseline measures of earnings inequality in the U.S. in 1980 and 2000. I focus on the college wage premium–the regression-adjusted difference in mean log earnings between college graduates and high school graduates–which is a common earnings inequality measure.[9] With 1980 and 2000 U.S. Census data, I estimate the regression equation:

$$w_{ij} = x_i\beta + e_{ij}, \tag{3}$$

where w_{ij} indicates the natural logarithm of person i's average weekly labor earnings observed in location j. The vector x_i includes indicators for education categories (high school drop-out, some post-secondary education but no degree, and college graduate), indicators for being female, black, and Hispanic, and an indicator for each year of poten-

[7]I was able to match all NLS-72 respondents to a high school CZ and more than 99 percent of them to a fourth follow-up CZ. The corresponding match rates for the NELS:88 were 97 percent and 99 percent. Unmatched respondents tend to be somewhat more educated but otherwise similar to others. Tables 8 through 10 show differences between respondents with and without usable location information.

[8]I assume a respondent has finished formal schooling if he or she does not return to formal schooling during the sample period.

[9]See Moretti (2010) for an example. Goldin and Katz (2008) document the relationship between the college wage premium and other measures of earnings inequality, such as the ratio of 90th percentile to 10th percentile earnings. Trends in inequality series are not always the same, but many measures of earnings inequality increased overall from the 1960s to the early 2000s. Lemieux (2006a) argues that the rising return to postsecondary education was a particularly important element of inequality trends in the U.S.

tial work experience from 0 to 39.[10]

Table 1 displays the wage regression results. Estimation in the 1980 Census yields a college premium of 0.415 log points, which rises dramatically to 0.61 log points in 2000. The wage return to some college and to high school graduation (the omitted education category) also increased. The increase in earnings inequality in the U.S. over this time period is the subject of much research, mostly investigating potential causes behind the trends (e.g., skill-biased technological change, changes in minimum wage laws and union activity, increasing international trade). Katz and Autor (1999) provide a survey. More recent contributions include Lemieux (2006b) and Autor, Katz, and Kearney (2007). My focus is on the contribution of location decisions to this rising earnings inequality.

My motivation for investigating the role of location choice in earnings inequality is the fact that firms pay similar-looking workers more in some locations than others. For example, firms in large cities offer higher wages and salaries than firms in smaller rural labor markets, a phenomenon labeled the "urban wage premium" and the focus of a large research literature.[11] It is very likely that locations paying relatively high wages to observationally similar workers have higher labor productivity than locations that pay less: otherwise, firms would move operations to areas with lower labor costs. In equilibrium, the higher wages in more productive places could be sustained if they are off-set by higher costs of living (a possibility I address below) or if workers have limited geographic mobility.

Below, I investigate whether young workers appear to sort differentially into more and less productive locations. In order to do so, I construct a measure of labor productivity for each location (CZ). To start, I estimate Equation (3) with nation-wide Census samples

[10]The regression sample includes men and women ages 18 to 64 who were not in group quarters. I include only those with unallocated work hours, work weeks, and labor earnings data. I replace top-coded earnings data with 1.5 times the top-code level ($75,000 in 1980 and $175,000 in 2000). I include only those who worked at least 40 weeks and at least 35 hours per week usually. The weekly wage variable is the annual wage and salary income divided by weeks worked. I weight the regression by the product of the person sample weight and the respondent's hours worked in the past year.

[11]See Glaeser and Maré (2001), Yankow (2006), Wheeler (2006), and Gould (2007) for analyses of the forces behind the urban wage premium.

(separately for 1980 and 2000) and collect the wage residual for each worker. I then take the average of workers' residuals in a given location j as a measure of "productivity" in that location. This is a measure of local average log wages conditional on education, sex, race, and potential experience, that is, by how much more or less firms in each location pay observationally similar workers.[12]

Although the wage residual measure does not perfectly capture local labor productivity, I interpret it as a proxy for local characteristics that increase overall labor productivity. These could be favorable physical features like a port or human capital externalities, for example. A very likely reason for wages–conditional on worker characteristics–to differ across locations is that different features of the locations induce a variety of local labor productivity levels. In the next subsection, I consider the potential for unobserved worker characteristics to drive the variation in local wage residuals.

In 1980 and 2000, I calculate the lowest average wage residuals for small CZs in the plains states. In 1980, the highest average wage residuals are in Alaska, consistent with the oil boom there in the 1970s. Manufacturing areas Gary, IN and Detroit, MI and large cities Chicago and Washington, D.C. also feature particularly high wages in 1980. In 2000, Alaska still features high wages, but at the top of the average wage residuals ranking is the San Jose, CA labor market, which includes Silicon Valley. Other large metropolitan areas dominate the top of the list in 2000. The relationship between CZ population and average wage residuals is positive in 1980 and increases in magnitude by 2000.

3.3 The effect of geographical sorting on unobservables

For my purposes, a potentially confounding reason for a local labor market to have high wages is differential selection of workers with high unobserved skills to some locations more than others. My framework implies that a location's productivity (wage) measure

[12] As described above, the Census data do not identify CZ of residence (or workplace) for workers, so I approximate CZ averages using county group averages. In practice, I calculate average wage residuals for each county group or PUMA. I then take a population-weighted average of PUMA averages that make up each CZ and call that the CZ productivity (average wage residual) measure.

gives a proxy for how much the wage would increase for a randomly-selected worker locating there. If my productivity measure is just capturing the local stock of unobserved skills, then it will not necessarily influence the earnings of a worker moving to that location.

While I cannot completely rule out an influence of differential selection on my productivity measure, I can provide some evidence that it is not likely to drive the results. When calculating the location-average wage residual, I control flexibly for important wage determinants including education, potential experience, race, ethnicity, and gender. A selection story would then need to rely on other wage determinants. Of course, there remains the possibility that other labor market skills are left in the residual and not measured in the Census data. So, I move to the NLS-72 and NELS:88 longitudinal data sets to explore the potential for unobserved skills to explain the location average wage residuals.

In Table 2, each column includes regression coefficients from a specification following Equation (3), but this time estimated using the longitudinal samples of early labor force participants (NLS-72 and NELS:88) rather than Census respondents. The dependent variable is the natural logarithm of annual earnings when respondents are around 25 years old. I include only those respondents working full-time who have completed their formal schooling.

The coefficients on the "College grad" variable of Table 2 replicate a large college premium in all specifications, and a much larger college premium among the later cohort. Columns 1 and 5 give the estimates most similar to Census estimates in Table 1. The smaller college premia in Table 2 reflect the much younger workers in the NCES samples: the college premium takes time to expand fully, but it is quite high even among young labor market participants. Columns 2 and 6 include as an additional regressor the labor productivity measure for the respondent's CZ of residence (described above). As expected, local labor productivity predicts higher individual earnings. This is reassuring given that the local productivity measure comes from Census data, and these earnings data come from NLS-72 and NELS:88.

Columns 3 and 7 test the extent to which some proxies for labor market skills unobserved in the Census can explain the relatively high wages in some CZs (that is, locational selection on unobserved skills). The first skill proxy is a transformed value from a test given to student respondents. I regressed reading and math scores on earnings later in life and from the regression coefficients created an index of predicted earnings given a student's test scores. Parent education is the maximum years of schooling between the respondent's mother and father, and family income is measured while the respondent is in secondary school. As expected, test scores and family income have positive partial correlations with earnings (conditional on these, the parent education relationship is very weak). Importantly, the relationship between the residence CZ productivity measure (average wage residual from the Census) and respondent wages remains strong after controlling for several proxies for skills unobserved in the Census.

Previous literature also supports the claim that some local labor markets exhibit higher wages for workers than others, for reasons other than differential sorting on unobserved worker traits. Moretti (2010) argues that wage differences across locations are driven by differences in labor demand conditions. As evidence, he shows that MSAs experiencing relatively large increases in their college graduate populations from 1980 to 2000 also experienced relatively large increases in their college wage premia. Higher local labor demand can explain higher local wages for a given resident without regard to his observed or unobserved characteristics.

My findings below imply that college graduates tend to locate in more productive local labor markets than high school graduates. Part of this locational difference comes from differences in NLS-72 and NELS:88 respondents' locations when they were in high school. Since high school students do not in general choose their location decisions, they are probably not likely to be driven by labor market sorting.[13]

When NLS-72 and NELS:88 respondents choose their own locations as adults, I also

[13] To the extent that parents pass their unobserved abilities to their children, there may be a link between unobserved abilities of high schoolers and average workers in their locations. However, I expect this connection to be somewhat weak, since it would rely on both an intergenerational correlation in skills (which is imperfect) and location sorting of parents (which is not clearly strong).

observe college graduates choosing more productive destinations than respondents with less education. This is *not* because some local labor markets offer high wages to college graduates–with relatively high unobserved skills as well–but not to other workers. Columns 4 and 8 of Table 2 show regression specifications that allow the relationship between the CZ wage productivity measure and individual workers' earnings to vary by education level. In both NLS-72 and NELS:88 samples, the effect is positive for all education categories. Indeed, more-productive locations offer an even larger wage increase to high school graduates than college graduates. So in comparison to college graduates, high school graduates appear to have at least as much to gain by moving to productive CZs, but they do so at a lower rate.

Another result I describe below is that college graduates tend to choose locations that feature both higher wages and higher costs of living, in about the same magnitude, relative to high school graduates. This finding appears inconsistent with the productivity measure simply capturing unobserved worker traits that would be rewarded equally in any location. If workers in high-paying CZs could earn the same wages for their transferable skills but face lower costs of living by moving, then they probably would. The fact that they tend not to implies that the local wage productivity measure proxies for a given worker's local wage in relation to offers in other locations.

3.4 Empirical findings about location choice and local wages

I use data on individual geographic mobility behavior from the 1970s (NLS-72) and 1990s (NELS:88), along with CZ average wage residuals, to discern how much earnings inequality is related to location choices. Tables 3 and 4 show summary statistics for the samples I use from the NLS-72 and NELS:88, respectively. NLS-72 provides a sample of approximately 13,690 and NELS:88 provides a sample of approximately 10,040 respondents with non-missing location information in 12th grade and the fourth follow-up surveys.[14] Both

[14]Following requirements for the use of restricted-access data, I round all unweighted sample sizes describing these data to the nearest ten. I follow this procedure throughout this paper.

tables show that migration out of one's 12th grade origin CZ is common and becoming more common over time. In the NELS:88 sample, about 34 percent of respondents lived at age 26 in a CZ other than their 12th grade CZ.[15]

Migration behavior varies substantially across education levels. Tables 3 and 4 both imply that college graduation is much more common among movers than stayers. Research consistently shows such a positive correlation between education and migration frequency (for example, Malamud and Wozniak (2009) argue that college graduation causes more frequent migration). In addition, the migration rate gap between higher and lower education groups increased between the 1970s cohort (NLS-72) and the 1990s cohort (NELS:88). The migration rate of college graduates increased from 41 percent to 49 percent, while the migration rate among high school graduates only increased from 17.5 percent to 18.4 percent.[16]

Women are slightly over-represented among movers in the 1970s cohort but not the 1990s cohort. Blacks and Hispanics are under-represented among movers in both cohorts. Following the well-known fact that age at first marriage is increasing, the percent of the sample that was married by age 26 in the 1970s is significantly higher than the associated percent in the 1990s, although the difference in the percent having children is not as large. Those who have children by age 26 are less geographically mobile than those without children, and the difference becomes large in the 1990s cohort.

Table 5 displays new findings about the relationship between educational attainment and geographic location. Each column presents results from a regression in which the observations correspond to respondents to either NLS-72 or NELS:88. The dependent variable is the wage-residual-based labor productivity measure of the respondent's CZ of residence: in Columns 1 and 2, this refers to the CZ where the respondent attended high school (12th grade).

[15] The use of sample weights is the reason that the given migration rate is not equal to the ratio of the mover sample size to the whole sample size.

[16] The migration rate among high school drop-outs in these cohorts fell. However, the sample here is not representative of all high school drop-outs: the sample frame is of high school seniors, so anyone who dropped out before 12th grade was not eligible to enter the sample.

In both cohorts, people who eventually attain more education tend to attend high school in more-productive locations. In addition, the difference between origin location productivity of college graduates and high school graduates increases markedly between the 1970s cohort and 1990s cohort. Relative to their peers who did not attend college, high school seniors in 1992 who went on to graduate from college lived in CZs where similar workers earned about 0.048 log points more on average. To the extent that productivity of one's high school location affects later earnings, it contributes to the difference in earnings between college graduates and those with less education. Such a contribution could be caused by the tendency of people to stay close to where they grew up and any productivity spill-over from local workers to local high schoolers. College graduates tend to benefit more–in terms of earnings–from their parents' location decisions than high school graduates do. Hence, location is a mechanism that dampens intergenerational earnings mobility.

Columns 3 and 4 of Table 5 focus on location choices made by the respondents themselves. The motivation behind these regression specifications is to assess how people with the same geographic origin (high school CZ) but with different characteristics make different location decisions. The estimation for these columns is by OLS where all the variables were subtracted by the origin-CZ mean in the sample (a within-origin-CZ estimator). Conditioning on high school location controls for any origin characteristic that might affect future location decisions. Examples are proximity to a large city, local institutions that make out-migration more and less attractive, and local labor market conditions like unemployment and job growth.

Conditional on where they attended high school, college graduates tend to move toward more productive locations than their less-educated peers. The difference in destination choice labor productivity (average log wage residuals) between college and high school graduates is 0.0266 log points in the 1970s and grows significantly to 0.0545 log points in the 1990s. The first thing to note is that college graduates choose more-productive labor markets. It was not clear ex ante that this would be the case, since college graduates may simply earn more in all locations and not be over-represented in highly productive locations. This finding complements others in the literature. Moretti (2010) shows that college graduates tend to live in areas with higher costs of living, which may be related to local productivity but also to amenability and local housing supply. Berry and Glaeser (2005) show that college graduates have been tending to cluster in cities with relatively high initial education levels and relatively high college wage premia. Neither study estimates migration behavior directly, but both assess location in cross-sections from the U.S. Census.

Hence, a substantial part of the earnings difference across education groups can be explained by more-educated workers choosing to live in locations where workers in general earn more. In addition, the growth in the college premium is related to growing educational differences in location choices. To show this more clearly, I collect estimates from Tables 1 and 5 to form Table 6. The first row–"Wage gap"–repeats the mean difference in log earnings of college and high school graduate workers, conditional on potential experience, race/ethnicity, and sex (estimated using 1980 and 2000 Census data). The second row repeats estimates of the mean difference in log earnings residuals between locations where college and high school graduates grew up. Four percent of the college wage premium in 1980 could be attributed to differences in origin locations of college and high school graduates. Even as the college wage premium increases from 1980 to 2000, the difference between origins of college and high school graduates can account for an increasing share of the college wage premium, becoming almost 8 percent in 2000.

Workers migrate away from their origin locations and thereby add to or subtract from the effect of their origin location on their earnings. Compared to high school graduates from the same origin CZ, college graduate movers in the 1970s cohort chose locations where similar workers earned on average 0.027 log points more. Since movers were about 30 percent of the cohort, such a difference implies that migration resulted in a difference in log earnings of about 0.008 between college and high school graduates, or almost 2 percent of the overall college wage premium. As with origin differences, migration differences increased over the sample period, even faster than the rise in the college wage premium. In the 1990s cohort, migration choices account for about 3.6 percent of the college wage premium.

Taken together, location at origin and subsequent migration decisions imply that college graduates in the 1970s cohort worked as young adults in CZs where similar workers were paid about 0.025 log points (0.017+0.008) more than high school graduates. The analogous difference in the 1990s cohort is 0.7, significantly larger. The share of the college wage premium accounted for by origin and early labor market migration effects grew from about 6 percent to about 11.5 percent. While these shares are somewhat modest, they indicate that a significant part of earnings inequality is associated with location differences, some chosen by workers and some chosen by their parents.

4 Differences in local costs of living by education level

The previous section demonstrated that differential origin locations and subsequent migration choices both contribute to earnings inequality and its rise in the U.S. College graduates grow up and choose to live in locations that have relatively large positive effects on their earnings. It is not yet clear why college graduates tend to cluster in high-paying locations. One possibility is that college graduates are displaying more ability to choose locations with robust labor markets and higher earnings: part of the return to schooling is this ability to increase earnings through migration. If this is the case, then policies that encourage strategic geographic mobility among the less-educated may yield more utility for them while decreasing earnings inequality.

However, local costs of living may very well adjust to compensate for differences in earnings across locations. Consider the economic model of multiple locations described in Section 2. The relatively high level of economic activity in a very productive location should induce relatively high local prices, as firms and workers bid up the price of local goods (e.g., land). This will tend to offset the wage gain to locating there. How much local prices adjust to local productivity differences depends on the elasticities of labor supply to a location and of local goods (e.g., housing supply). Hence, it is an empirical question.

In this section, I focus on differences between college and high school graduates in their local costs of living. To the extent that college graduates choose locations with high costs of living, they reduce the consumption or utility impact of their locational advantage in nominal wages. I find that this is the case: college graduates choose locations with costs of living that offset most of their location wage advantage. An implication of this is that levels and increases in earnings inequality in the U.S. are larger than levels and increases in cost-of-living-adjusted earnings. This finding, complementing similar recent evidence in Moretti (2010), implies that U.S. earnings inequality–though large in magnitude–is higher than inequality in purchasing power.

4.1 Data description

There is no comprehensive set of cost of living indices that are consistent over time, specific to locations in the U.S., cover the whole country, and include multiple types of local prices (housing and other prices).[17] However, U.S. Census data include housing prices that cover the entire country and are consistent over time. I therefore use local housing prices to proxy local costs of living. Housing accounts for a large share of household expenditures and is clearly the most important location-specific price category.

The method I use to construct location-specific housing price indices follows Moretti (2010). I use household-level data from the 1980 and 2000 U.S. Censuses (Ruggles, et al. 2010). I select only households in 2- or 3-bedroom housing units (not in group quarters). For the results reported here, I select only renters and use the variable "monthly gross rent" as the baseline housing price measure (using only values that were not allocated).[18]

[17] Moretti (2010) provides a helpful discussion of availability of local price data. The Bureau of Labor Statistics publishes local price indices including non-housing goods, but they do not allow comparisons across cities and only include a limited number of large cities. The ACCRA Cost of Living Index includes non-housing consumption goods prices for a large sample of cities over time, but the sample sizes are somewhat small.

[18] I also estimated specifications using a housing cost index from owners' self-reported house values. The results are very similar, since location-specific rental and housing prices move very closely together, as

This rental variable includes utility costs. For top-coded values, I impute 1.3 times the top-code ($999 in 1980 and $9,999 in 2000).

I calculate the average 2- or 3-bedroom unit's monthly rent for each county group (or PUMA) identified in each Census year, weighting by the Census household weights. Using the same procedure described above for Census average wage residuals, I estimate Commuting Zone average rental costs as population-weighted averages of the county group average housing costs that make up each CZ. I then convert these CZ-average housing rental prices to CZ-specific housing cost indices. I adjust current dollars to 1980 dollars using the CPI-U. Local costs of living are a weighted average of prices of local goods and traded goods that have a single price across all locations. To weight local prices in overall costs of living, I adopt the housing share in expenditures used to calculate the CPI-U. These are the "relative importance" of housing expenditures in the CPI-U closest to the dates of my data (U.S. Department of Labor). Using this method, about 60 percent of costs of living are common across locations, but about 40 percent vary across households in different locations.

4.2 Empirical findings about location choice and local costs of living

Table 7 displays results from regressions that describe the costs of living where respondents in NLS-72 and NELS:88 reside. The dependent variable in columns 1 and 2 is the log of the rental cost index where the students attended 12th grade (their origins). Students who go on to obtain more education tend to originate in locations with higher costs of living. In addition, the gap has grown between the 1970s and the 1990s: the gap between college and high school graduate costs of living increased from 0.0319 log points to 0.0645 log points.

Columns 3 and 4 describe costs of living in destination locations. As with the regressions above describing destination average wage residuals, these are within-origin estimates: all the variables were subtracted by the origin-CZ mean in the sample. So, the

expected.

coefficient on "College grad" of 0.0365 in column 3 implies that college graduates choose to reside in a location with costs of living 0.0365 log points higher than high school graduates *from the same origin CZ*. Hence, these education-related differences in destination choice do not depend on origin locations, which affect for example whether very productive potential destinations are nearby. Again, the difference between college and high school graduates in destination costs of living increase between the 1970s and 1990s cohorts, to 0.059 log points in the 1990s.

The previous section showed that college graduates choose locations where workers earn more on average than locations that high school graduates choose. This section shows that higher costs of living compensate for those higher earnings in college graduates' destinations. Comparing the magnitude of average earnings and cost of living differences implies that much of the location-related earnings gain from college graduates' locations is wiped out by higher costs of living. For example, the NELS:88 college graduates chose to live in destinations where workers earn about 5.5 percent more but also where costs of living are about 6 percent more.

This does not mean that they are worse off, however. Inter-city differences in nominal wages *fully* adjusted for local prices probably misrepresent inter-city utility differences somewhat (Dumond, Hirsch, and Macpherson 1999). Residents of high-cost cities should substitute away from locally-priced (expensive) goods, relative to residents of low-cost cities. This dampens the difference between their utilities relative to the difference between the costs they face for consuming a fixed bundle of goods. Hence, the difference between local price indices (including the same goods with the same weights) between college graduates and high school graduates probably overstates the difference in the costs of achieving the same utility level.

I interpret the findings of Tables 5 and 7 to imply that college graduates choose more productive labor markets than their high-school graduate peers. College graduates earn a larger location premium in wages as a result, but they do not expand their consumption opportunities, since relatively high costs of living take away approximately all of their location wage premium. So, college graduates are increasingly clustering, by origin and choice, in local labor markets featuring both relatively high wages and relatively high costs of housing. Firms would not be willing to pay relatively high wages and land costs in these cities unless they also feature relatively productive amenities. These could be human capital externalities, physical attributes like a port, access to markets, or spillovers from other industries.

5 Conclusion

This paper provides new empirical evidence about how workers' locations affect measurements of earnings inequality in the United States. I find that a significant and rising proportion of the college wage premium is due to college graduates living in and moving toward higher-paying locations than high school graduates. Furthermore, I find differences in the local costs of living where college and high school graduates choose to live: college graduates choose more expensive locations. The differences between college and high school graduates in location wages and location-specific price indices are approximately equal. From this, I infer that college graduates choose to live in more economically productive labor markets (higher wages and higher costs of living) than do workers with less education.

One implication is that nominal earnings inequality measures (not accounting for local costs of living) are higher than inequality in purchasing power (which depend on costs of living). This complements similar findings in Moretti (2010). I find that origin locations are particularly important in determining the wages and costs of living where workers end up living as adults. However, migration behavior also moves college graduates to relatively high-wage and high-cost-of-living locations.

My findings imply that college graduates on average do not parlay their higher education into better migration decisions in terms of finding locations where higher real earnings (and more consumption) are available, relative to workers with less education.

So, my findings *do not* imply that less-educated workers are engaging in too little or poorly-targeted migration decisions. Encouraging workers with relatively low education to migrate would not necessarily increase their utility or reduce inequality in purchasing power.

However, it is possible that college graduates still gain in other ways from their relative concentration in more productive locations. More productive locations may feature more job offers, which could help reduce unemployment spells. A similar benefit would accrue if more productive locations are less susceptible to economic downturns. Assessment of such effects is outside the scope of this paper but appears to be a promising line of future research.

References

[1] Autor, David H. and David Dorn (2008) "Inequality and Specialization: The Growth of Low-Skill Service Jobs in the United States" MIT Working Paper.

[2] Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008) "Trends in U.S. Wage Inequality: Revising and Revisionists" *Review of Economics and Statistics* 90(2) 300-323.

[3] Berry, Christopher and Edward L. Glaeser (2005) "The Divergence of Human Capital Levels across Cities" *Regional Science* 84(3) 407-444.

[4] Black, Dan, Natalia Kolesnikova, and Lowell Taylor (2009) "Earnings Functions When Wages and Prices Vary by Location" *Journal of Labor Economics* 27(1) 21-47.

[5] Curtin, Thomas R., Steven J. Ingels, Shiying Wu, and Ruth Heuer (2002) *National Education Longitudinal Study of 1988: Base-Year to Fourth Follow-up Data File User's Manual* (NCES 2002-323). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

[6] Dumond, J. Michael, Barry T. Hirsch, and David A. Macpherson (1999) "Wage Differentials across Labor Markets and Workers: Does Cost of Living Matter?" *Economic Inquiry* 37(4) 577-598.

[7] Glaeser, Edward L. and David C. Maré (2001) "Cities and Skills" *Journal of Labor Economics* 19(2) 316-342.

[8] Goldin, Claudia and Lawrence F. Katz (2008) *The Race between Education and Technology* (Cambridge: Harvard UP).

[9] Gould, Eric D. (2007) "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium" *Review of Economic Studies* 74, 477-506.

[10] Katz, Lawrence F. and David H. Autor (1999) "Changes in the Wage Structure and Earnings Inequality" Chapter 26 in Orley Ashenfelter and David Card, eds. *Handbook of Labor Economics*, Volume 3A (Amsterdam: Elsevier) 1463-1555.

[11] Lemieux, Thomas (2006a) "Postsecondary Education and Increased Wage Inequality" *American Economic Review* 96(2) 195-199.

[12] Lemieux, Thomas (2006b) "Increased Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review* 96(3) 461-498.

[13] Malamud, Ofer and Abigail Wozniak (2009) "The Impact of College Education on Geographic Mobility: Identifying Education Using Multiple Components of Vietnam Draft Risk" mimeo.

[14] McHenry, Peter (2010) "The Geographic Distribution of Human Capital: Measurement of Contributing Mechanisms" College of William and Mary.

[15] Mincer, Jacob (1978) "Family Migration Decisions" *Journal of Political Economy* 86(5) 749-773.

[16] Moretti, Enrico (2010) “Real Wage Inequality” UC Berkeley working paper.

[17] Roback, Jennifer (1982) “Wages, Rents, and the Quality of Life” *Journal of Political Economy* 90(6) 1257-1278.

[18] Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

[19] Solon, Gary (1999) “Intergenerational Mobility in the Labor Market” Chapter 29 in Orley Ashenfelter and David Card, eds. *Handbook of Labor Economics*, Volume 3A (Amsterdam: Elsevier) 1761-1800.

[20] Tolbert, Charles M. and Molly Sizer (1996) “U.S. Commuting Zones and Labor Market Areas. A 1990 Update.” Economic Research Service Staff Paper No. 9614.

[21] U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88/94 and NELS:88/2000). Restricted use files.

[22] U.S. Department of Education, National Center for Education Statistics, National Longitudinal Study of the High School Class of 1972 (NLS-72).

[23] U.S. Department of Labor, Bureau of Labor Statistics, *CPI Detailed Reports*, various years, Table 1: Consumer Price Index for All Urban Consumers: U.S. city average, by expenditure category and commodity and service group.

[24] Wheeler, Christopher H. (2006) “Cities and the Growth of Wages among Young Workers: Evidence from the NLSY” *Journal of Urban Economics* 60, 162-184.

[25] Yankow, Jeffrey J. (2006) “Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium” *Journal of Urban Economics* 60, 139-161.

6 Figures and Tables

Table 1: College Wage Premium in 1980 and 2000

VARIABLES	(1) 1980	(2) 2000
College grad	.415***	.61***
	(.00533)	(.00494)
Some PSE	.139***	.191***
	(.00234)	(.00157)
Ed< HS	-.218***	-.298***
	(.00305)	(.0027)
Female	-.44***	-.304***
	(.0051)	(.00204)
Black	-.156***	-.126***
	(.0106)	(.00513)
Hispanic	-.136***	-.143***
	(.0112)	(.00451)
Constant	4.99***	5.66***
	(.00658)	(.00679)
Observations	2376448	2692450
R-squared	0.351	0.349

***p<0.01 **p<0.05 *p<0.1. US Census data from 1980 and 2000. Dependent variable is log weekly wage. Regressions also include separate intercept for each year of potential experience. High school graduate with no college education is the omitted education category. Regressions weighted by product of person weight and annual hours worked. Standard errors clustered at county group level.

Table 2: Determinants of Log Annual Earnings

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	NLS-72 (1979 earnings)				NELS:88 (1999 earnings)			
College grad	.244***	.229***	.173***	.226***	.479***	.425***	.33***	.431***
	(.0179)	(.0172)	(.019)	(.0169)	(.0321)	(.0304)	(.0303)	(.0305)
Some PSE	.0977***	.0854***	.0635***	.0831***	.207***	.19***	.156***	.2***
	(.0128)	(.0123)	(.0125)	(.0123)	(.0322)	(.029)	(.0272)	(.0282)
Ed< HS	-.00562	-.019	-.0246	-.0213	-.0619	-.11	-.0727	-.102
	(.0982)	(.098)	(.0999)	(.091)	(.0913)	(.0898)	(.0864)	(.0862)
Yrs Experience	.0357***	.0372***	.0392***	.0373***	.047***	.0513***	.0498***	.0518***
	(.00378)	(.00349)	(.00351)	(.00346)	(.00692)	(.00663)	(.00676)	(.00665)
Female	-.343***	-.345***	-.337***	-.346***	-.28***	-.278***	-.257***	-.278***
	(.0114)	(.0109)	(.0109)	(.011)	(.0219)	(.0205)	(.0189)	(.0203)
Black	-.0895***	-.088***	-.0264	-.0876***	-.0792**	-.122***	-.0497	-.124***
	(.016)	(.0148)	(.0173)	(.0149)	(.038)	(.0311)	(.0325)	(.0309)
Hispanic	-.045	-.0409	.00815	-.0418	-.0502	-.0863	-.0312	-.0855
	(.0399)	(.0303)	(.0293)	(.03)	(.0744)	(.0542)	(.0431)	(.052)
Residence CZ		.782***	.73***	.904***		.973***	.931***	.909***
productivity		(.0652)	(.0681)	(.108)		(.0993)	(.102)	(.136)
Test Index × 100			.00908***				.0139***	
			(.00163)				(.00231)	
Parent ed			.00216				-.00352	
			(.00188)				(.00533)	
Log Family inc			.0475***				.0461***	
			(.00884)				(.0146)	
College				-.415***				-.0751
× CZ prod.				(.144)				(.21)
Some PSE				-.0322				.22
× CZ prod.				(.109)				(.207)
Ed< HS				-.0316				-.53
× CZ prod.				(1.05)				(.534)
Constant	9.37***	9.38***	.475	9.39***	9.97***	10***	-4.36*	10***
	(.0312)	(.0277)	(1.51)	(.0274)	(.0436)	(.0408)	(2.3)	(.0409)
Observations	8770	8770	8770	8770	5160	5160	5160	5160
R-squared	0.199	0.230	0.239	0.232	0.142	0.187	0.204	0.188

***p<0.01 **p<0.05 *p<0.1. Dependent variable is individual's log annual earnings (1979 for NLS-72, 1999 for NELS:88). Full-time workers only. Sample weights used. Standard errors clustered at residence CZ level. The test score index is a prediction of adult earnings conditional on a student's math and reading test scores. Family income is a measure from when respondent was in secondary school. Parent education is the higher of mother's and father's years of schooling. Specifications with test scores, family incomes, and parent education also include indicators for missing values, which were replaced by predictions based on other respondent characteristics.

Table 3: Characteristics of NLS-72 Respondents, by Move Status

	(1)	(2)	(3)	(4)	(5)	(6)
	All		Movers		Stayers	
	[$N \approx 13690$]		[$N \approx 3740$]		[$N \approx 9950$]	
			[$Rate = .27$]		[$Rate = .73$]	
Variable	Mean	SE	Mean	SE	Mean	SE
HS Dropout	.005	.001	.003	.001	.005	.001
HS Grad	.21	.004	.136	.006	.238	.005
Some PSE	.516	.005	.457	.009	.539	.006
College Grad	.269	.004	.404	.009	.218	.005
Female	.499	.005	.524	.009	.489	.006
Black	.09	.002	.062	.004	.1	.003
Hispanic	.037	.002	.022	.002	.043	.002
Married by Age 26	.636	.005	.672	.008	.622	.005
Any Kids by Age 26	.392	.005	.357	.009	.406	.005

Sample sizes rounded to the nearest 10 due to confidentiality restrictions. Move status is determined by residence in a commuting zone (CZ) at age 26 that is different from the CZ of the last high school attended. HS, PSE, and SE refer to high school, post-secondary education, and standard error, respectively.

Table 4: Characteristics of NELS:88 Respondents, by Move Status

	(1)	(2)	(3)	(4)	(5)	(6)
	All		Movers		Stayers	
	$[N \approx 10040]$		$[N \approx 3580]$		$[N \approx 6460]$	
			$[Rate = .338]$		$[Rate = .662]$	
Variable	Mean	SE	Mean	SE	Mean	SE
HS Dropout	.008	.002	.004	.001	.011	.003
HS Grad	.154	.007	.084	.006	.189	.009
Some PSE	.487	.009	.405	.012	.529	.012
College Grad	.346	.008	.505	.012	.265	.009
Female	.502	.009	.502	.012	.503	.012
Black	.118	.01	.079	.009	.137	.014
Hispanic	.098	.005	.058	.005	.119	.008
Married by Age 26	.459	.009	.452	.012	.462	.012
Any Kids by Age 26	.346	.009	.233	.01	.403	.012

Sample sizes rounded to the nearest 10 due to confidentiality restrictions. Move status is determined by residence in a commuting zone (CZ) at age 26 that is different from the CZ of the last high school attended. HS, PSE, and SE refer to high school, post-secondary education, and standard error, respectively.

Table 5: Individual Traits Predicting Labor Productivity (Avg. Wage Residuals) in Residence CZs

	(1)	(2)	(3)	(4)
	Dep. Var.: Origin CZ Productivity (OLS)		Dep. Var. Destination CZ Productivity (Within-Origin Estimates)	
VARIABLES	NLS-72	NELS:88	NLS-72	NELS:88
College grad	.0167***	.0483***	.0266***	.0545***
	(.00453)	(.0122)	(.00567)	(.00992)
Some PSE	.0128***	.0242***	.0198***	.0173*
	(.00327)	(.0067)	(.00623)	(.0096)
Ed< HS	.0257**	.0242	.00936	.00153
	(.0125)	(.0166)	(.0391)	(.0311)
Female	.00172	-.00438	.00518	-.00503
	(.00199)	(.00417)	(.00355)	(.00417)
Black	-.00691	.0486***	.0395***	.0275**
	(.00998)	(.0117)	(.00777)	(.011)
Hispanic	-.0126	.0368	.0201*	.000802
	(.0214)	(.0253)	(.0113)	(.012)
Constant	-.0231***	-.067***	3.08e-11	-1.08e-10
	(.00888)	(.0111)	(1.76e-10)	(2.40e-10)
Observations	13690	10030	3740	3580
R-squared	0.005	0.037	0.016	0.039

***p<0.01 **p<0.05 *p<0.1. Dependent variable is residence CZ average log wage residual (wage-based productivity measure). Origin is CZ of high school attended as a senior. Destination is CZ of residence at fourth follow-up (age 26). Sample weights used. Standard errors clustered at origin CZ level.

Table 6: Locational Components of College Log(Earnings) Gap, 1980 and 2000

	(1)	(2)	(3)	(4)
	1980		2000	
	Value	% of Wage Gap	Value	% of Wage Gap
Wage gap	.415	100	.61	100
Origin effect on wages	.017	4.03	.048	7.91
Migration effect (Δ_m) on wages	.027		.055	
Migration rate (P_m)	.306		.398	
$\Delta_m \times P_m$	.008	1.93	.022	3.61

Table 7: Individual Traits Predicting Cost of Living (Log Rent Index) in Residence CZs

	(1)	(2)	(3)	(4)
	Dep. Var.: Origin CZ Log(Rent index) (OLS)		Dep. Var. Destination CZ Log(Rent index) (Within-Origin Estimates)	
VARIABLES	NLS-72	NELS:88	NLS-72	NELS:88
College grad	.0319***	.0645***	.0365***	.059***
	(.00587)	(.0128)	(.00773)	(.00942)
Some PSE	.0296***	.0362***	.0313***	.0235**
	(.00528)	(.00755)	(.00762)	(.00942)
Ed< HS	.0224	.0386**	.0496	.0051
	(.0185)	(.0188)	(.0391)	(.0287)
Female	.00254	-.00497	.00308	-.0106**
	(.00263)	(.00447)	(.00466)	(.00481)
Black	-.0182	.0437***	.0298***	.025**
	(.0121)	(.0142)	(.0108)	(.0123)
Hispanic	.0364	.0907***	.0217*	.0049
	(.0279)	(.0337)	(.0112)	(.0128)
Constant	.0898***	.0706***	-9.58e-10	-8.51e-10
	(.0113)	(.012)	(5.95e-10)	(7.89e-10)
Observations	13690	10030	3740	3580
R-squared	0.015	0.061	0.014	0.032

***p<0.01 **p<0.05 *p<0.1. Dependent variable is residence CZ log rental price index (cost-of-living measure). Origin is CZ of high school attended as a senior. Destination is CZ of residence at fourth follow-up (age 26). Sample weights used. Standard errors clustered at origin CZ level.

7 Data appendix: Identifying locations in NLS-72 and NELS:88

I was not always able to identify residence Commuting Zones (CZs) for all respondents to the NLS-72 and NELS:88. Tables 8 to 10 compare characteristics of respondents for which I do and do not have CZ identified. I was able to identify origin locations for all NLS-72 respondents, so there is no table comparing NLS-72 respondents with identified and not identified origins. I drop respondents with missing CZ from the analysis.

Table 8: Characteristics of NELS:88 Respondents, by High School Location Data Status

	(1)	(2)	(3)	(4)
	CZ identified [$N \approx 10100$]		CZ not identified [$N \approx 300$]	
Variable	Mean	SE	Mean	SE
HS Dropout	.008	.002	.016	.011
College Grad	.347	.008	.351	.046
Test Index	10.022	.001	10.025	.003
Parents Ed	14.044	.034	13.857	.23
Parents Income	44943	546	47919	3088
Female	.502	.009	.501	.049
Asian	.042	.003	.111	.027
Hispanic	.098	.005	.138	.028
Black	.117	.01	.043	.017
White	.731	.01	.696	.041
Married by Age 26	.459	.009	.427	.05
Any Kids by Age 26	.345	.008	.266	.041

CZ means Commuting Zone, a location definition. The test score index is a prediction of adult earnings conditional on a student's math and reading test scores. Family income is a measure from when respondent was in secondary school. Parent education is the higher of mother's and father's years of schooling.

Table 9: Characteristics of NLS-72 Respondents, by Fourth Follow-up Location Data Status

	(1)	(2)	(3)	(4)
	CZ identified $[N \approx 13690]$		CZ not identified $[N \approx 160]$	
Variable	Mean	SE	Mean	SE
HS Dropout	.005	.001	.007	.007
College Grad	.269	.004	.329	.041
Test Index	9.304	0	9.312	.003
Parents Ed	12.866	.022	13.468	.211
Parents Income	10809	47	11414	462
Female	.5	.005	.423	.043
Asian	.011	.001	.014	.014
Hispanic	.037	.002	.029	.012
Black	.09	.002	.084	.022
White	.841	.003	.839	.03
Married by Age 26	.635	.005	.636	.042
Any Kids by Age 26	.392	.005	.348	.042

CZ means Commuting Zone, a location definition. The test score index is a prediction of adult earnings conditional on a student's math and reading test scores. Family income is a measure from when respondent was in secondary school. Parent education is the higher of mother's and father's years of schooling.

Table 10: Characteristics of NELS:88 Respondents, by Final Follow-up Location Data Status

	(1)	(2)	(3)	(4)
	CZ identified $[N \approx 10330]$		CZ not identified $[N \approx 70]$	
Variable	Mean	SE	Mean	SE
HS Dropout	.008	.002	0	0
College Grad	.346	.008	.521	.086
Test Index	10.022	.001	10.037	.011
Parents Ed	14.034	.034	14.602	.299
Parents Income	44967	538	55177	8247
Female	.503	.009	.372	.08
Asian	.044	.003	.047	.022
Hispanic	.1	.005	.026	.014
Black	.115	.01	.073	.037
White	.729	.01	.847	.046
Married by Age 26	.458	.009	.556	.085
Any Kids by Age 26	.343	.008	.349	.087

CZ means Commuting Zone, a location definition. The test score index is a prediction of adult earnings conditional on a student's math and reading test scores. Family income is a measure from when respondent was in secondary school. Parent education is the higher of mother's and father's years of schooling.