THE ROLE OF LOCATION IN EVALUATING RACIAL WAGE DISPARITY

DAN A. BLACK, NATALIA KOLESNIKOVA, SETH G. SANDERS, AND LOWELL J. TAYLOR

Abstract. A standard object of empirical analysis in labor economics is a modified Mincer wage function in which an individual's log wage is specified to be a function of education, experience, and an indicator variable identifying race. The hope is that estimates can be informative about the impact of minority status on labor market success. Here we examine the theoretical basis for this regression in a context in which individuals live and work in different locations. Our theory calls into question the conventional approach, which implicitly assumes that the "race wage gap" is a single parameter; instead, the race wage gap is predicted to vary by location. With this insight in mind, we reevaluate evidence about the black-white wage disparity in the United States.

JEL: J31, J71, R23.

Keywords: wage regressions, racial wage disparity, theory of local labor markets.

Introduction

In hundreds of studies social scientists have examined the role of minority status in wage determination by estimating variants of the Mincer earnings function,

$$\ln(w_i) = \gamma(x) + \beta_1 R_i + \beta_2 E_i + \epsilon_i; \tag{1}$$

the expected log wage of individual i is specified to be a function of an indicator variable for minority demographic status R_i (e.g., race, immigrant status, ethnicity, or gender) and education E_i (or an alternative measure of human capital) and $\gamma(x)$ represent covariates typically added such as age or experience. An statistically significant estimated negative value of β_1 is taken as an indication of wage disparity that adversely affects members of the minority population. Such disparity might arise from disparate treatment in the labor market, or if minority individuals have lower levels of productivity than others with similar levels of measured human capital.

Date: March 2009. Author affiliations: Black, University of Chicago and NORC; Kolesnikova, Federal Reserve Bank of St. Louis; Sanders, Duke University; and Taylor, Carnegie Mellon University. The views expressed are ours and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors. Helpful comments on earlier drafts were provided by David Card, Enrico Moretti, and seminar participants at UC Berkeley.

Here we study the the properties of the estimates of regression (1) when individuals live in locations that have differing prices, e.g., different local wages and housing prices. We are interested in both of the possibilities mentioned above—that wage gaps are due to discrimination or due to unmeasured differences in human capital. Either way we find that in a standard model of local labor markets, β_1 is unlikely to be a single parameter; instead, race wage gaps are predicted to differ systematically vary by location.

With this observation in mind, we turn to two empirical exercises. First, we conduct an analysis of black-white wage gaps in the United States from 1940 through 2000—work that roughly parallels the well-known work of Smith and Welch (1989). Second, we conduct an analysis of black-white wage gaps, as well as wage gaps between Hispanics and non-Hispanic whites, in which we control for cognitive test scores (AFQT scores), as in Neal and Johnson's (1996) important paper. In both cases, we discover that a theoretically-motivated approach that accounts for racial differences in location yields results that differ substantially from approaches that implicitly ignore location.

Our paper proceeds in four additional sections:

The first section examines a standard model of urban differences in prices, and in that context demonstrate that the race wage gap—whether generated by human capital differences or discrimination—is a constant across different local labor markets if and only if preferences are homothetic. This latter case is special, and seems unlikely to generally pertain. We show, moreover, that even if preferences are homothetic, common implementation practices with Mincer earnings regressions may be problematic.

The second section presents empirical evidence for the importance of this idea for assessing the convergence of black-white wage gaps in the U.S. form 1940 through 2000.

The third section presents empirical evidence for the importance of location in the evaluation of race and ethnicity wage gaps that control for cognitive skills.

Section four provides a discussion.

1. Race Wage Gaps in a Multiple-Location Model

Workers supply labor in local labor markets, and across those markets there are often substantial differences in wages and housing prices. Theoretical reasoning in the urban/regional economic literature, in the pioneering work of Haurin (1980) and Roback (1982) and in many papers that followed, suggests that observed location-specific price differences can generally be understood to be the consequences of differences in locations' attractiveness or location-specific differences in productivity. Our goal is to determine what these models have to say about racial wage disparities in local labor markets.

1.1. A Simple Economic Analysis

We begin with a population in which individuals belong to one of two racial groups: a minority group, indicated by $R = 1$, and a majority group, indicated by $R = 0$. These people live in one of n cities, and consume two goods: a non-housing good that has a price 1 in every location, and housing, which has a price that varies across cities. We designate the rental price of housing p_j per unit $(j = 1, \dots n)$. Wages also differ across location, and we want to allow for the possibility of race-based differences in wages: individuals from group $R = 0$ earn wage w_j^0 in city j, while those from group 1 earn w_j^1.

For simplicity, we assume that all workers supply one unit of labor, regardless of where they work. We assume also that all individuals have the same preferences. Finally, we assume that there is costless migration between locations. Let the expenditure function for workers of each group ($R = 0$ or 1) living in city j be $e_j^R = e(p_j, u_j^R)$. The key equilibrium condition is that workers of both groups must be indifferent over their city of residence; for individuals in each race group, utility u_j^R is the same in each city. Therefore we can drop the subscript j on utility, and note that equilibrium entails

$$e(p_j, u^0) = w_j^0 \quad \text{and} \quad e(p_j, u^1) = w_j^1 \quad \text{for} \quad j = 1, \dots, n. \tag{2}$$

Note that while u_j^0 and u_j^1 are the same in each city, utility might differ *between* demographic groups if their earnings differ. As we have noted, this latter outcome is possible if the groups differ in the levels of human capital or if there is racial discrimination. With this in mind, we consider a "disparity index" in location j, which we define to be the ratio of the wage for the minority group 1 relative to the wage of the majority group 0,

$$I_j = \frac{w_j^1}{w_j^0} = \frac{e(p_j, u^1)}{e(p_j, u^0)}. \tag{3}$$

This disparity index will be less one if the minority group is disadvantaged or greater than one if the minority group is advantaged. Importantly, in general this ratio is seen to depend on the housing price p_j.

When is the disparity index independent of location-specific price variation? First, note that if preferences are such that individuals' expenditure functions takes a "separable" form $e(p,u) = \psi(p)f(u)$, the disparity index in location j is $I_j = \dfrac{\psi(p_j)f(u^1)}{\psi(p_j)f(u^0)} = \dfrac{f(u^1)}{f(u^0)}$, which does not depend on local prices. Second, and more importantly, note that the converse is true. The proof is simple: Let $I_j = g(u^0, u^1)$, so that the index in location j does not depend on that location's prices. Without loss of generality we can take $u^0 = 1$, $u^1 = u$. Then $I_j = \dfrac{e(p_j,u)}{e(p_j,1)} = g(u,1)$, so we can write $e(p_j,u) = e(p_j,1)\cdot g(u,1)$. Setting $\psi(p) \equiv e(p,1)$ and $f(u) \equiv g(u,1)$ we find that the expenditure function has the form $e(p,u) = \psi(p)f(u)$.

A familiar result from price theory is that the expenditure function takes the form $e(p,u) = f(u)\psi(p)$ if and only if preferences are homothetic. We thus have a key proposition: *In an equilibrium model of local labor markets, the racial disparity index is the same across locations if and only if preferences are homothetic.*

As long as preferences are homothetic, it proves quite easy to relate our theory back to the familiar Mincer wage regression (1). Under homotheticity, the form of the expenditure function is $w = \psi(p)f(u)$. Using the logarithmic form of this latter equation, we notice that for a person i living in city j and who has race R_i we expect

$$\ln(w_{ij}) = \ln(\psi(p_j)) + \ln(f(u^{R_i})), \tag{4}$$

where R_i indicates the individual's race ($R = 0, 1$). As the $\psi(p_j)$ is independent of utility, local wage levels are seen to vary with prices, but log racial wage disparity is a constant that is invariant with regard to location. Thus we can let $b_0^j \equiv \ln(\psi(p_j)$ and $\beta_1 \equiv \ln(f(u^{R_i}))$, and we have the structural relationship

$$\ln(w_{ij}) = \beta_0^j + \beta_1 R_i, \tag{5}$$

where β_1 the penalty (or premium) to minority status. If we add a random error term to equation (5) we have a simplified version of the familiar Mincer regression (1). Suppose, in addition, that the achievable level of utility is increasing in education, in the following way: For an individual with education E_i we can write $f(u) = \bar{u}e^{\gamma(x)+\beta_1 R_i+\beta_2 E_i+\epsilon_i;}$.[1] Then we have the equation

$$\ln(w_{ij}) = \beta_0^j + \gamma(x_i) + \beta_1 R_i + \beta_2 E_i + \epsilon_i. \tag{6}$$

This is the same as the familiar equation (1) with one very important exception: If housing prices and wage levels differ by location, we must include a fixed effect for each local labor market ($j = 1, \cdots, n$) under study.

In fact in the U.S. there is large variability in housing prices. Thus, even if one has taken great care to look construct a sample of minority and majority individuals who have similar levels of human capital, so that the simple regression (5) will do (i.e., one is simply looking for the proportional racial difference in wages), if one fails to include an indicator variable for each local labor market, the OLS estimate of β will be inconsistent, except in the unusual case in which the distribution of race is the same across cities.[2]

In short, our theory leads us back to the traditional specification of the earnings functions *if preferences are homothetic,* and with the important caveat that we must include city fixed effects unless there is little price variation between cities (or unless the distribution of workers by race is identical across locations). Unfortunately, homotheticity is a strong restriction, implying that for all goods the income elasticity is equal to one. In fact, a large literature suggests that for many goods the income elasticity is very different than one for many goods, including, importantly, housing.[3] If preferences are *not* homothetic, the race wage gap cannot be the same across cities, i.e., there will be no single parameter β_1 in equation (6), but rather a parameter β_1^j for each city j.

1.2. An Example with Location-Specific Productivity Differences

[1]Mincer (1974) provides the theoretical justification for this form. There is an important caveat. In Black, Kolesnikova and Taylor (2009) we find that β_2, the "return to education," will differ by location if preferences are not homothetic. Here, though, we *are* assuming homotheticity, and in that case returns are the same for each location.

[2]As we show below, in the U.S. there are very large differences in the location patterns of black and white men.

[3]Estimating an income elasticity of demand on housing is difficult (see Olsen, 1987, for a discussion). One widely cited study, Rosen (1985), reports an income elasticity of demand on housing of 0.76. Harmon (1988) reviews a large number of studies and concludes that an estimate of the income elasticity of demand of 0.7 may be appropriate for most applications.

To illustrate the mechanism at work in our model we turn to a specific example. Let us suppose for the moment that differences in location stem from differences in *productivity* in the locations. While there has been much work on possible causes of these productivity differences (e.g., see Acemoglu, 1996, Glaeser and Mare, 2001, and other work on agglomeration), we remain agnostic as to the source of the variation. Whatever the source of such variation in productivity, the city with higher productivity will have higher wages and in consequence will typically have higher housing prices.

Let us suppose also that minority workers have a lower level of unobserved human capital than majority-group workers. We ask how racial wage disparities will differ across these locations.

Continue to let u^1 and u^0 be utility levels, respectively, of minority and majority workers. Given that minority workers have a lower level of human capital, and thus within each city lower wages, their utility will also be lower; $u^1 < u^0$. The disparity index in a given city with a housing price p is $I = \dfrac{e(p,u^1)}{e(p,u^0)}$.

We want to know how this index in a low-price, low-productivity city compares to the index in a higher-price city. We conduct this thought experiment by evaluating the derivative of the disparity index with respect to the housing price:

$$\frac{\partial I}{\partial p} = \frac{1}{e(p,u^0)^2}\left[e(p,u^0)\frac{\partial e(p,u^1)}{\partial p} - e(p,u^1)\frac{\partial e(p,u^0)}{\partial p}\right]. \tag{7}$$

With a bit of algebraic manipulation we can rewrite (7):

$$\frac{\partial I}{\partial p} = \frac{e(p,u^1)}{pe(p,u^0)}\left[\frac{p}{e(p,u^1)}\frac{\partial e(p,u^1)}{\partial p} - \frac{p}{e(p,u^0)}\frac{\partial e(p,u^0)}{\partial p}\right]. \tag{8}$$

Now Shephard's lemma indicates that the derivative of the expenditure function with respect to p is the demand for housing. So (8) can in turn can be written in terms of the budget shares of housing for minority workers and majority workers, respectively s_H^1 and s_H^0:

$$\frac{\partial I}{\partial p} = \frac{e(p,u^1)(s_H^1 - s_H^0)}{pe(p,u^0)}. \tag{9}$$

This latter expression is positive if minority workers allocate a higher share of their income to housing than do their majority counterparts. Given that minority workers have relatively lower income, this amounts to the assumption that the income elasticity of housing is less than one—and assumption broadly supported by the literature. We conclude that for this case

$$\frac{\partial I}{\partial p} > 0. \tag{10}$$

Given that cities with relatively high productivity are also cities with higher housing prices in this example, we expect that the wage disparity index will be *higher* in high-productivity cities than in low-productivity cities. This means that for a disadvantaged minority, the disparity index will be closer to 1; the proportional nominal wage gap will be *lower* in the high-productivity city.

It is quite easy to explain the logic of this proposition. Suppose individuals live in one of two locations—to take a concrete example, say Memphis and Chicago in 1940—and suppose that in each location black workers (the minority in this example) earn less than their white counterparts because of differences in human capital. Suppose further that all workers are more productive in Chicago, owing perhaps to Chicago's extensive railroad system and industrial agglomerates. In equilibrium we expect Chicago to have higher wages than Memphis and also to have a relatively higher housing price. What about the black-white wage gap in the two cities? Given that the elasticity of demand for housing is less than one, the relatively high housing price in Chicago places a greater burden on the (poorer) black workers than the (richer) white workers. Thus if both black and white workers are indifferent between living in Memphis and Chicago, as they must be in equilibrium, black workers will require a larger "Chicago wage premium" than will white workers; the proportional gap between black and white wages will be *lower* in Chicago than in Memphis. This latter outcome, of course, is what our derivation (10) shows.

It is important to notice that the same conclusion follows if the minority wage gap is instead generated by labor market discrimination rather than human capital differences. Notice, first of all, that under our assumption of costless mobility, a discriminated-against black worker will be willing to live in either city, Memphis and Chicago, only if utility is the same in the two locations. Thus the equilibrium condition (2) continues to hold, as do our subsequent derivations, leading to (10). Again, the resulting wage disparity must be smaller in Chicago than in Memphis. Intuitively, the *utility cost* must be the same in the two places, and this can happen only if the proportional wage gap is smaller in the location with higher housing prices.

As we have mentioned, there is another mechanism for generating price differentials across locations—differentials in location-specific amenities. In general, it is difficult to know how these might relate to differences in the black-white wage gap across location, especially since the two groups might value the amenities differently.[4] Here we simply note that our general observation continues to hold: in equilibrium, if prices differ by location, so too will the black-white wage gap.

It is easy to see why the insights of our model might be important for general inferences about black-white economic disparity. Suppose, to continue our example, that a black individual moves from Memphis to Chicago. In our model, his wage increases but his utility is unchanged. A social scientist measuring disparity with a traditional wage regression will show disparity declining even though *real* disparity—disparity measured in terms of welfare—clearly has not changed at all. Is this issue important in practice? We turn to that question in the next section.

2. The Importance of Location for Evaluating the Black-White Wage Gap

A generation of labor economists is now familiar with the basic picture presented in Smith and Welch's seminal 1989 paper, "Black Economic Progress After Myrdal." Using data from the U.S. Census, Smith and Welch calculate that the black-white gap in weekly wages is 0.57 in 1940, 0.45 in 1950, 0.42 in 1960, 0.36 in 1970 and 0.27 in 1980.

In this section update these basic facts about the black-white gap by evaluating also results in 1990 and 2000, and we proceed with an additional contribution: we evaluate the role of location in drawing inferences about the black-white wage gap.

2.1. Estimates of the Black-White Wage Gap, 1940-2000 Census Data

Like Smith and Welch, and many other authors, we here evaluate the black-white wage gap using public use samples from the U.S. Decennial Census. There are substantial advantages of these data for this purpose. First, it provides us with an opportunity to examine the economic progress of African Americans relative to whites over a long period, 1940

[4]For example, one amenity might be the size of one's own racial community.

through 2000, using data from instruments that are similar both in terms of content and mode of administration. Second, the data provide extremely large samples, and therefore allow for precise estimates.

As many other scholars have noted, though, there are unfortunately rather severe limitations with the Census data in regard to the variables available. We do however have data on key economic outcome variables like earned income and labor supply, along with race, age, and education, and we have some information on location of residence. With these variables will proceed as best we can. Even though the variables are quite limited, we are able to establish quite convincingly our central point—that treatment of location is very important in the estimation and interpretation of the decline in black-white disparity in U.S. labor markets over the past six decades.

For our analysis, we restrict attention to men.[5] We begin by dividing respondents on the basis of race—black and non-Hispanic white—and exclude other racial/ethnic groups. We are interested in wages earned by "prime aged" full-time working men, so we restrict attention to men aged 25-55 who worked at least 27 weeks in the previous year.[6] In our analysis we make use *age*, which we have in 31 discrete categories (individual years, 25 though 55 inclusive), *education*, which we have in 10 categories ("no schooling or kindergarten only" through "more than a bachelor's degree"), and *location*, which we have for several hundred unique localities.[7]

Our primary focus is on measuring black-white wage disparity, conditional on at least some observable characteristics. Given our discussion above, it is clear that we do not view black-white wage disparity as a single parameter. At a minimum the disparity is likely to differ by location, and it may also vary by level of education and by age. So we will be typically estimating a weighted mean of observed disparities. The key of course is the construction of sensible weights given available data.

[5]Female labor markets are equally interesting, and we intend to evaluate black-white wage gaps among women in future work.

[6]We exclude unpaid family workers, military personnel, the self-employed, and those employed in agriculture. See the data appendix for more detail. In general we pattern our exclusion rules after Smith and Welch, although there are some substantial differences. The data appendix also outlines how we construct the key wage variable.

[7]The data appendix discusses our location variables. These vary somewhat over the 60 years of analysis.

To give the simplest possible example, suppose we are interested in conditioning only on age. We can proceed as follows. Let b index black individuals and w index white individuals, and let x_i be the exact year age of individual i. Let y_i be the log wage of individual i, and let $E(y_{b,i}|x)$ be the expected value of the log wage of that (black) individual given that he is age x. Our interest then is in

$$\Delta = \sum_{x=25}^{55} [E(y_{b,i}|x) - E(y_{w,i}|x)] f_b(x), \tag{11}$$

where $f_b(x)$ is p.d.f. of age among black workers. The idea of looking at the object $[E(y_{b,i}|x) - E(y_{w,i}|x)]$ is of course that $E(y_{w,i}|x)$ provides a missing counterfactual to the question: what would be the expected log wage of a black worker age x if he were treated in the labor market as a white worker with that same age.[8] Then by averaging difference over the age distribution of *black* workers we are looking at the "average treatment effect on the *treated*."

Our theoretical reasoning suggests that the Δ will differ by location. If we are interested in the "average treatment effect" over all locations, we can follow an approach comparable to that given in (11) but now let x index a location-age cell, e.g., one cell will be *men aged 31 residing in Houston.* Notice now that in the Census data there will be thousands of such cells, which again we index with x. Now we have

$$\Delta = \sum_{x=0}^{N} [E(y_{b,i}|x) - E(y_{w,i}|x)] f_b(x), \tag{12}$$

where N is the number of age-location cells.

Finally, of course, there is a tradition in race wage regressions of controlling also for schooling. Given that education in our data is categorized in discrete cells (as discussed in the appendix), we needn't change the basic non-parametric approach. In this instance we simply let x index a location-age-education cell, e.g., *high-school educated men aged 31 in Houston*, and now let $f_b(x)$ represent the distribution of the black population over these cells.

[8]Notice that the "treatment" here is *not* the absence of disparate treatment (if any) by employers; being treated as a white person in society includes also other facets, e.g., pre-market outcomes that can lead to disparities in human capital.

In principle, one could directly estimate equation (12) by calculating the conditional means at each point in the distribution of covariates. As a practical matter, we implement an estimation procedure that returns us to the traditional regression framework. In particular, let $\hat{\Delta}$ be the non-parametric matching estimator based the direct approach. As alternative suppose we simply use weighted OLS on the basic regression

$$y_i = \beta_0 + \beta_1 R_i + \epsilon_i, \tag{13}$$

and designate that estimator $\hat{\beta}_1$. With a bit of algebra it is possible to establish that $\hat{\Delta} \equiv \hat{\beta}_1$ if the weights are constructed appropriately.

The first step in constructing the weights is to realize that the Census data themselves come with weights that allow one to mimic the U.S. population. In the appendix we describe our treatment of missing data. Our approach is to assume that data are, conditional on the age-race-education-race cell, missing at random. We thereby construct new weights; for an individual in a particular cell x^0 the weight, adjusted for missing data, is $w_1(x^0)$. Now consider the conditional "probability of being black" for a that particular cell:

$$p(x^0) = Pr(Black|x = x^0). \tag{14}$$

Having calculated this probability for each cell, we may define a new final set of weights, say $w_2(x^0)$, as follows:

$$w_2(x^0) = \begin{cases} w_1(x^0) & \text{if the worker is black, and} \\ w_1(x^0)\frac{p(x^0)}{1-p(x^0)} & \text{if the worker is white.} \end{cases} \tag{15}$$

Notice that if there is a white worker who is not matched to a black worker at all in the data ($p(x^0) = 0$), that individual is simply dropped from the analysis, and, similarly, if his characteristics quite dissimilar from blacks in the sample he will be given low weight. Conversely, white individuals who have characteristics that are more typical of the black individuals in the sample are weighted more highly. Intuitively, our re-weighting scheme forces the distribution of covariates in the sample of whites to be identical to the distribution of covariates in the sample of blacks. In the matching context, this is often referred to as "inverse probability weighting".[9]

It is important to keep in mind that the estimate of the average treatment effect contains not only the impact of the "observables" but also the impact of "unobservables." Thus, for example, if we implement our estimator by matching on all available observables (age, location, and education), we are still leaving out important ways in which black and white workers differ in the labor market. For example, Black, Haviland, Sanders, and Taylor (2006) document that black men choose college majors that are systematically less lucrative than those chosen by white men. Because the Census does not contain information on college major (i.e., it is an unobservable), we are unable to condition on this variable. Similarly, Neal (2006) documents large differences in the cognitive test scores of African Americans relative to whites. Again, the lack of, say, test scores in the Census means that the impacts of such scores are imbedded in the unobservables and their correlation with observable measures.

Finally, we remind the reader that we are analyzing wages of men who work 27 weeks a year or more. While the wages of working individuals are indeed important, so are the issues concerning racial differences in labor force nonparticipation. As for this matter, we note three well-known facts: First, nonparticipation rates of African American men are higher than the corresponding nonparticipation rates of whites. Second, nonparticipation rates are inversely correlated with education, and presumably nonparticipation also varies with unobservable skills as well. Third, nonparticipation rates have been growing over time.[10] Chandra (2000) gives an excellent review of the issues involved. Here we ignore these issues, and focus instead on the role of location for understanding the racial wage gap among those who are working.[11]

---

[9] See Hirano, Imbens, and Ridder (2003) and DiNardo, Fortin, and Lemieux (1996) for an extended discussion. See Black, Haviland, Sanders, and Taylor (2006, 2008) for applications to discrete data.

[10] Furthermore, much evidence (e.g., Black, Daniel, and Sanders, 2002, and Autor and Duggan, 2003) suggests this nonparticipation due to disability is quite sensitive to prevailing economic opportunities, particularly for the low skilled. In addition, of course, the increased incarceration of black males, noted by Western (2006) and others, also makes the use of observed wages problematic.

[11] In Black, Kolesnikova, Sanders, Taylor and Wessel (2009) we undertake additional explorations of the role of location using broader economic outcomes, e.g., income.

Table 1 gives our initial set of results, which use log weekly wage as the dependent variable in our key regression (13). We estimate this regression using weighted OLS with weights given in (15). In column (1) we report the outcome in which we match on age only. There is, of course, a compelling reason to match on age, since productivity is related to age, and since age is, from the perspective of labor market participant, exogenous. Having done so, we find that a log black-white gap of an astonishing -0.74 in 1940. This gap declines to a still-substantial -0.31 in 1980 and to -0.30 in 2000.[12] An important feature of our estimates is convergence; the log wage gap declines by approximately 44 log points over the period of study.

It is tempting to think of the raw wage gaps (adjusted for age only) given in column (1) as money-metric measure of the welfare disadvantage to being black in the American labor market. Given our theoretical discussion above, though, such an interpretation is unlikely to be correct. In our model such an inference can be made only after matching black and white individuals *within* labor markets. One can think about the average impact of race across these markets. Column (3) of Table 1 reports the resulting estimates of this latter sort. There are substantial differences in the inferences we draw. We see, for example, that within local labor markets the 1940 log wage gap is now "only" -0.66, instead of -0.74. Apparently in 1940 blacks disproportionately resided in labor markets that had relatively low wages for all workers. By 2000, though, the racial distribution of residence had changed substantially, and in consequence an log wage gap that accounts for location is larger in absolute value than one that does not, -0.36 vs. -0.30. Taking this approach we find that convergence of 30 log points, not 44 log points. Put another way, failure to treat location properly leads us to overestimate black-white wage convergence by nearly 50 percent.

As we have noted, it is common in the literature to condition on both age and education when evaluating wage gaps. The idea is to try to sort out how much of the race "treatment effect" is due to differences in years of formal schooling acquired by workers. We thus conduct our exercise matching on age and education in column (2) and on age, education, and location in column (4). In this case also, inferences about convergence are greatly

[12]For small values the log wage gap is approximately equal to the percentage wage gap. This approximation is not very good, though, for gaps as large as those we observe here. The percentage wage gaps implied by our estimates for 1940 through 2000 are, respectively, -0.52, -0.40, -0.40, -0.36, -0.27, -0.24, and -0.26.

affected by matching on location: In the absence of conditioning on location, the gap declines by 37 log points, but within location the gap is dropping by 24 log points. Failure to condition on location again leads us to overestimate black-white wage convergence by approximately 50 percent.

In preparing estimates for Table 1 we used as the dependent variable, log *weekly* wage, the same variable examined by Smith and Welch. It turns out that there are differences between black and white workers in the number of hours typically worked per work, so we might alternatively evaluate *hourly* wages in constructing racial wage gaps. Results of this latter analysis are reported in Table 2. The general inferences are similar to one would draw from Table 1. We substantially overestimate black-white wage convergence over the 1940-2000 period if we fail to account for shifting patterns of location.

What are the shifting patterns of residence that have such an important impact when we estimate black-white wage gaps? Table 3 provides the basic answer. In that table we report the results of the following exercise: We begin by calculated the extent to which black men disproportionately reside in the South. We do this by constructing an index equal to the ratio of "the fraction of black men aged 25-55 living in the South" to "the fraction of white men aged 25-55 living in the South." Notice that this index will equal 1 if black and white prime-aged men are equally likely to live in the South. We construct this same index for urban residency. The table indicates that in 1940 black men were very heavily over-represented in the South—the index is nearly 3—and very under-represented in urban areas. By 2000, the over-representation in Southern residence weakened substantially, and black men were disproportionately likely to live in urban areas. These large changes in residential patterns make a substantial difference when thinking about black-white wage inequality.

Interested readers can see trends in the black-white wage gap within several key cities in Table 4. This table shows the 14 cities with the largest black population concentrations of black men in 1940. One notable feature highlighted in the Table is that Southern cities generally had the largest wage gaps in 1940. To the extent that this variation represents an equilibrium phenomenon, this might suggest that Southern cities had relatively lower productivity or lower levels of "consumption amenities" than their non-Southern counterparts.

A second notable feature is that in many cities—including New York and Philadelphia, and all of the Midwestern cities—there was virtually no narrowing of the black-white wage gap after 1970. Indeed, in Cleveland, Chicago, and Detroit, there was little change in the black-white gap over the entire 5 decade span, 1950–2000.

Our central message, in any event, is that the central message of black-white wage convergence is greatly overstated—overstated by approximately 50 percent—when we fail to control properly for location.

3. The Importance of Location for Evaluating the Relationship Between Race, Cognitive Ability and Earnings

In an highly influential paper, Neal and Johnson (1996) provide an important critique of the standard modified Mincer wage regression as applied to the analysis of race wage disparities. They note that (1) blacks and whites typically have different levels of human capital, even conditioned on observed years of schooling, and, in any event (2) completed years of schooling is an endogenous choice variable that will depend on any number of factors, including the quality of schooling to which a young person has been exposed. They argue, therefore, in favor of an alternative approach in which the wage regression includes a measure of cognitive ability, measured while an individual is still quite young (i.e., while he or she is a teenager), rather than the more traditional "years of schooling" variable.[13] In such an empirical exercise, estimated black-white gaps are found to be quite small in absolute value. Indeed, Bollinger (2003), in his reanalysis of Neal and Johnson's data (which looks at the role of measurement error inherent in any test score) summarizes by suggesting that human capital "attainment at age 18 may explain all of the gross differences in wages between blacks and whites" (p. 583).

Now our theory, presented in Section 1, suggests that in regressions of the sort estimated by Neal and Johnson, black-white gaps are likely to vary across location except under one *best case* (or, more precisely, one *easiest case*) scenario: when preferences are homothetic. Even in this best case, though, our theory suggests that one must include location fixed effects in the wage regression. With this in mind, we conduct here a reexamination of the

[13] O'Neill (1990) also includes a test score measure in her wage regressions in the analysis of black-white wage gaps.

central results of Neal and Johnson (1996) and Bollinger (2003), using, as did these authors, data from the National Longitudinal Study of Youth, 1979, and updating also to the 1997 cohort. We are able to look at black-white wage gaps, and also wage gaps between Hispanics and non-Hispanic whites.

We begin by considering the simplest wage equation of Neal and Johnson, and Bollinger:

$$\ln(w_i) = \alpha_0 + \alpha_1 \text{Age}_i + \beta_B R_i + \beta_H H_i + \epsilon_i, \tag{16}$$

where $\ln(w_i)$ is the natural logarithm of the respondent's wage on the last job, Age_i is the respondent i's age in months, R_i in a race indicator variable equal to 1 if the respondent is black, and H_i is an indicator variable of Hispanic ethnicity. $\alpha_0, \alpha_1, \beta_B$, and β_H are parameters to be estimated, and ϵ_i is the regression error. We use data from the 1979 Cohort of the National Longitudinal Survey of Youth (NLSY) for the year 1990. We report the estimates of β_B and β_H this equation separately for men and women in column (1) of Table 5.

Our estimates and sample differ somewhat from those of Neal and Johnson (1996) and Bollinger (2003) because we use the full NLSY 1979 data (while the Neal and Johnson data are limited to those born in 1962 and 1964) but the estimates are very similar. For men, the estimated values of both β_B and β_H are quite large in absolute value, approximately -0.25 and -0.11 respectively. Black men and Hispanic men earn less than their similarly-aged non-Hispanic white counterparts. For women, blacks earn less than whites, but Hispanics are seen to have similar earnings to non-Hispanic whites.

In column (2), we repeat the analysis but, motivated by our theoretical reasoning above, we now allow for the error term to have a location specific intercept, that is, for each location l we have $\epsilon_{il} = \eta_l + e_i$. The inclusion of the location fixed effects increases the absolute value of the estimated disparity coefficients for blacks, and especially for Hispanics. Apparently, blacks and Hispanics disproportionately live in locations in which the wages of non-Hispanic whites are relatively high.

We turn next to the regressions in which we include a standardized measure of cognitive skills. In this case we use the results of the Armed Forces Qualification Test (AFQT), as do Neal and Johnson. Our specification now, like Bollinger's, is

$$\ln(w_i) = \gamma_0 + \gamma_1 \mathrm{Age}_i + \gamma_2 \mathrm{AFQT} + \beta_B R_i + \beta_H H_i + \epsilon_i; \tag{17}$$

the AFQT score is entered linearly. We report the estimates of this equation, without the location fixed effects, in column (3). Results are again quite similar to those of Neal and Johnson (1996) and Bollinger (2003). For men, we cannot reject the hypothesis that non-Hispanic men earn the same as non-Hispanic white men. Indeed, the point estimate for β_H is positive. For black men, the coefficient is remains negative and statistically significant, but is reduced substantially in absolute value, from -0.25 to -0.06. For women, we obtain a surprising positive, statistically significant, value for the Hispanic indicator variable. We also find that black women earn substantially more than white women, and the difference is statistically significant.[14]

Finally, in column (4) we estimate the model (17) with location fixed effects. For men, the magnitude of the coefficient on black-white disparity more than doubles in absolute value, from -0.06 to -0.13. For Hispanic men, the disparity coefficient similarly declines substantially; the changes is approximately -0.06 (and the coefficient switches sign, but remains statistically insignificant). For women, both the black and the Hispanic coefficients remain positive, but they are reduced substantially in magnitude (and are no longer statistically significant).

Comparison of columns (3) and (4) indicate, in summary, that location plays a large role in determining the level of the earnings gaps. There are good reasons to be suspicious of the results for women, given concerns about selection into the labor force—selection which might vary by race and ethnicity (see Neal, 2004). For black men, though, the implications are striking. We obviously *cannot* conclude, on the basis of available evidence that all of the wage gap is due to cognitive skill differences between the races that develop at young ages. Even conditioning on the AFQT skill measure, black men are found to earn approximately 13 percent less than their white counterparts in their same labor markets.

In Table 6, we repeat our empirical exercise using data from the 2006 survey for the 1997 cohort. For men, results are, if anything, even more dramatic. Comparing columns (3) and

[14]This differs somewhat from Neal and Johnson. We are investigating further.

(4) we see that the estimated black-white gap declines from -0.11 to -0.18, i.e, by -0.07, and the estimated gap between Hispanics and non-Hispanic whites declines from 0.03 to -0.06, i.e., by -0.09. As for women, the inclusion of location fixed effects does not matter much in the estimation of black-white gaps, but has the same very large impact observed for men in analyzing gaps between Hispanics and non-Hispanic whites.[15]

We view the results in Tables 5 and 6 as presenting compelling evidence of the importance of conditioning location when comparing the earnings of groups with differing locations.

4. Concluding Remarks

We have described a simple model in which prices vary across location. Our model suggests that race wage gaps—or, for that matter, wage gaps between any sub-populations—are likely to vary across locations in systematic ways. The theory indicates that at a minimum researchers should include location fixed effects in any estimation of minority wage gaps. Empirical evaluation shows that inferences one draws concerning important labor market phenomena are substantially affected when we follow this latter proscription.

[15]Interestingly, in comparison to the estimates for the 1979 cohort, there is a sharp decline in the magnitude of the coefficient on the AFQT test. There are many possible explanation for this decline, we will mention but four. First, the cohorts are much different in age at this analysis: 25 to 32 for the 1979 and 22 to 26 for the 1997. Participation patterns may differ dramatically between those ages, as well as by the type of jobs held by respondents. Second, for the 1979 cohort, the Department of Defense provided the test scores using item response theory. For the 1997 cohort, however, the Department of Defense has been unwilling to provide the norming, so the test was normed by staff at the Center for Human Resources Research at Ohio State without benefit of the individual test items. It is possible that the inability to use individual test items may have substantially reduced the accuracy of the test norming. Third, the Department of Defence has moved to computer assisted testing for the AFQT; it is possible that the new AFQT is less predictive of civilian labor market success. Finally, it is, of course, possible that the economic rewards to cognitive skills of young workers have declined.

Data Appendix for Census Analysis

All the Census data for this paper are taken from IPUMS for the 1940 through 2000 Censuses; see Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2008) for details. The IPUMS represents an integrated data set of the Public Use Micro Samples (PUMS) that were released in each of these Censuses.

Respondents were asked about their earnings in the previous year, the number of weeks worked that year, and, at least for the 1980 to 2000 Censuses, the usual hours worked that year. Baum-Snow and Neal (2006), however, document systematic biases that differ by race and sex in responses to hours worked. We thus report results for both weekly earnings and hourly earnings.

Our goal is to provide an analysis similar to that of Smith and Welch (1989). Toward that end, we make many data-use decisions that parallel theirs, though there are differences that we outline here. Like Smith and Welch we restrict our analysis to workers who work at least 27 weeks. As for age restrictions, Smith and Welch consider men aged 16 to 60. We are concerned about the growth of enrollment in high school and college, and we do not want to worry about decisions of "early retirement," so we limit our analysis to men 25 to 55 inclusive. To deal with the issue of schooling, Smith and Welch drop men from their sample who are enrolled in school if they work less than 50 weeks a year. Given our age restrictions, we find that adjustment to be unnecessary. Smith and Welch exclude unpaid family workers, military personnel, and the self-employed who are not in agricultural. We also exclude unpaid family workers, military personnel, the self-employed, and *all* agricultural workers. Because of the increased mechanization of agricultural production in the U.S., there has been a dramatic reduction in farm labor and a corresponding increase in the size of farms; farming has become quite capital intensive. It is therefore difficult to separate the returns to capital from the returns to labor. We exclude wage-and-salary agricultural workers because payments to workers often involve payments in-kind, which makes the valuation of the wage paid difficult. Of course, the exclusion of agricultural workers has little effect in 2000, but represents a major exclusion for the early years. Excluding self-employed agricultural workers has the added advantage of rendering the 1940 Census compatible with subsequent Decennial Censuses, as the Census did not ask for farm earnings in 1940.

We also follow Smith and Welch in limiting the sample to workers whose reported weekly earnings meet a minimum limit on weekly wages and an upper limit. The adopted limits are as follows:

	Lower limit of weekly earnings	Upper limit of weekly earnings
1940	1.50	125
1950	3.25	250
1960	6.25	625
1970	10.00	1,250
1980	19.80	1,875
1990	35.00	3,000
2000	45.00	4,000

The limits for 1940 to 1980 are taken from Smith Welch (p. 522, footnote 5). For 1990 and 2000, we indexed the 1980 values to the CPI and rounded up.

An important concern with the Census data is *item nonresponse.* Respondents occasionally choose not to answer questions about their age, race, ethnicity, or education level. More frequently, respondents omit answers to questions about hours worked or earnings. Our approach is to drop respondents who do not answer questions about age, race, Hispanic status, education, or earnings. We do, however, increase the weights on other respondents with identical ages, race, and education levels to reflect the missing data by using inverse probability weighting. To be precise, we estimate the probably of a nonresponse, or

$$\Pr(NR = 1 | X = x^0) = F(X^0),$$

where x indexes the age-race-education-location cell, and then we construct weights,w_1 ,

$$w_1(x^0) \equiv \frac{w_0}{1 - F(x^0)},$$

where w_0 are the initial Census weights. Thus, if half the people in the age-race-education-location cell do not respond to their earnings or hours worked questions, the responders within the cell have their weights doubled (see Wooldridge (2007)). This procedure implicitly assumes that data are, conditional on the age-race-education-location cell, missing at random. Because we condition on age, race, education, and location, this procedure also replicates the Census joint distribution of the age-race-education-location variables.

We also face two additional important issues about the data: the measurement of education and the measurement of location. The measurement of education in the Censuses presents problems for at least three reasons. First, in 1990, the Census Bureau reworked the education question to account for highest degree for those with a college education and some categorical data for lower levels of education. For instance, in 1990 the Census asked: *How much school has this person COMPLETED? Fill ONE circle for the highest level COMPLETED or degree RECEIVED. If currently enrolled, mark the level of previous grade attended or highest degree received.* Response options were:

◦ No school completed ◦ Nursery school
◦ Kindergarten
◦ 1st, 2nd, 3rd, or 4th grade
◦ 5th, 6th, 7th, or 8th grade
◦ 9th grade
◦ 10th grade
◦ 11th grade
◦ 12th grade, NO DIPLOMA
◦ HIGH SCHOOL GRADUATE - high school DIPLOMA or the equivalent (e.g., GED)
◦ Some college but no degree
◦ Associate degree in college - Occupational program
◦ Associate degree in college - Academic program
◦ Bachelor's degree (For example: BA, AB, BS)
◦ Master's degree (For example: MA, MS, MEng, MEd, MSW, MBA)
◦ Professional school degree (For example: MD, DDS, DVM, LLB, JD)
◦ Doctorate degree (For example: PhD, EdD)

Prior to 1990, the Census asked instead about "years of schooling." For instance, in 1980 the Census asked: *What is the highest grade (or year) of regular school this person has ever attended? Fill one circle. If now attending school, mark grade person is in. If high school was finished by equivalency test (GED), mark "12".* Response options for *highest grade attended* were ◦ Never attended school, ◦ Nursery school, ◦ Kindergarten, and these further options:

Elementary through high school (grade or year)

1	2	3	4	5	6	7	8	9	10	11	12
○	○	○	○	○	○	○	○	○	○	○	○

College (academic years)

1	2	3	4	5	6	7	8 or more
○	○	○	○	○	○	○	○

Drawing consistent inferences with schooling data drawn these two ways is in principle quite simple if the two types of questions have a similar structure of measurement error. Unfortunately, for the 1990 Census, Black, Sanders, and Taylor (2003) document that the education questions exhibit significant measurement error and that the degree of measurement error is correlated with race. Moreover, there was a dramatic increase in the educational attainment of Americans over the period. For instance, in 1940 88 percent of blacks and 64 percent of whites between the ages of 25 and 60 did not have a high school education, and only 2 percent of blacks and 8 percent of whites had a bachelor's degree or better. By 2000, only 9 percent of blacks and 5 percent of whites did not have a high school degree while fully 33 percent of whites and over 19 percent of blacks had a bachelor's degree or better.

In our regression analysis we treat education in a non-parametric way, and given the available data, we use the following ten education categories: no formal education or kindergarten only, 1 to 4 years, 5 to 8 years, 9 years, 10 years, 11 years, 12 years, some college but no bachelor's degree, bachelor's degree, and more than a bachelor's degree.

Finally, there is the issue of the measurement of location. Because of the growth in cities and changes in disclosure policy, the identification of metropolitan statistical areas (MSAs) varies over time. In 1960 (the first public use micro sample that the Census Bureau released), the only geography identified was State of residence. As a result, we cannot conduct the same location analysis of interest to us with the 1960 data; we use only an urban indicator interacted with an indicator for state of residence. In 1940, 1950, and 1970 through 2000, we use MSA of residence for those respondents living in a MSA. For those respondents not living in an identified MSA, we use an indicator for state of residence. Hence, we exploit the geographical variation that is generally available to us. There are, however, a few additional noteworthy limitations:

First, residents of some current MSA's are not separately identified in the early censuses, but are so identified subsequently. For example, in the 1940, Orlando residents are treated as individuals living in "rural" Florida, but in later years are broken out as part of an Orlando MSA. Similarly, Las Vegas is identified only starting in 1970. There are a host of smaller towns that are only identified in later years. Moreover, MSAs can be created from regions that were previously a part of different MSAs. This is a particular problem in the densely populated areas of the east and west coasts. Finally, for areas that are only identified as "rural" we may be mixing residents from very different areas of a given state. For example, this designation mixes residents of the desert areas of Southern California with residents of rural Northern California, who may face very different labor markets and price levels.

For the 1940 through 1970 Censuses, we use a one-percent sample of respondents, and from 1980 to 2000, we use the five-percent sample, which, along with population growth, provides much larger sample sizes and much more precise estimates. Prior to 1980, the Census did not release the five-percent sample so we do not have access the extremely large sample sizes that the five-percent sample affords. Finally, we note one important data limitation with the 1950 Census. In 1950, only the "sample line" respondents were asked about education and earnings by the Census Bureau. Hence, only about 3.3 percent of the population was given these questions. Thus, estimates from the 1950 Census are considerably less precise than estimates from even the 1940 Census.

In 1960 and 1970, the Census asked only for hours of work and weeks of work on intervals. To impute the actual levels, we took information from the 1980 Census and calculated the average weeks (or average hours) conditional on the being in the relevant category. In the table on the next page we provide the values that we used for the imputations.

Interval	Imputed weeks	Interval	Imputed hours
1-13 weeks	1.1	1-14 hours	8.57
14-26 weeks	21.4	15-29 hours	21.95
27-39 weeks	33.3	30-34 hours	30.64
40-47 weeks	43.4	35-39 hours	36.35
48-49 weeks	48.3	40 hours	40
50-52 weeks	51.8	41-48 hours	45.46
		49-59 hours	51.41
		60 or more hours	67.02

Prior to 1980, the Census did not ask the usual hours worked so we used hours last week as a proxy. In 1980, conditional on both reports being positive, the correlation is only 0.61. While quite low, this correlation is not materially different than those found in validation studies; see Barron, Black, and Berger (1997) for a discussion.

Table 1. Black-White Gaps in Log Weekly Wage

	(1)	(2)	(3)	(4)
Age	Yes	Yes	Yes	Yes
Education	No	Yes	No	Yes
Location	No	No	Yes	Yes
1940	-0.741	-0.584	-0.662	-0.495
	(0.0119)	(0.0093)	(0.0081)	(0.0086)
1950	-0.511	-0.400	-0.485	-0.365
	(0.0139)	(0.0110)	(0.0108)	(0.0125)
1960	-0.510	-0.372	-0.489	-0.366
	(0.0070)	(0.0041)	(0.0063)	(0.0050)
1970	-0.447	-0.315	-0.448	-0.329
	(0.0138)	(0.0057)	(0.0054)	(0.0047)
1980	-0.308	-0.238	-0.332	-0.256
	(0.0125)	(0.0049)	(0.0031)	(0.0022)
1990	-0.281	-0.212	-0.323	-0.248
	(0.0054)	(0.0054)	(0.0027)	(0.0022)
2000	-0.301	-0.214	-0.358	-0.257
	(0.0093)	(0.0050)	(0.0033)	(0.0024)

Source: Authors' calculations, 1940 to 2000 I-PUMS. Dependent variable is the logarithm of weekly earnings. See the data appendix for details.

Table 2. Black-White Gaps in Hourly Wage

	(1)	(2)	(3)	(4)
Age	Yes	Yes	Yes	Yes
Education	No	Yes	No	Yes
Location	No	No	Yes	Yes
1940	-0.757	-0.636	-0.656	-0.522
	(0.0112)	(0.0092)	(0.0091)	(0.0098)
1950	-0.471	-0.374	-0.439	-0.341
	(0.0150)	(0.0133)	(0.0122)	(0.0145)
1960	-0.441	-0.318	-0.415	-0.308
	(0.0064)	(0.0043)	(0.0060)	(0.0054)
1970	-0.386	-0.260	-0.387	-0.276
	(0.0129)	(0.0061)	(0.0052)	(0.0050)
1980	-0.237	-0.170	-0.260	-0.189
	(0.0114)	(0.0046)	(0.0032)	(0.0024)
1990	-0.227	-0.162	-0.265	-0.196
	(0.0046)	(0.0046)	(0.0025)	(0.0021)
2000	-0.244	-0.165	-0.294	-0.202
	(0.0086)	(0.0044)	(0.0030)	(0.0023)

Source: Authors' calculations, 1940 to 2000 I-PUMS. Dependent variable is the logarithm of the hourly wage. See the data appendix for details.

Table 3. Location Indices for Black Men

	(1)	(2)
Year	**Southern Residency**	**Urban Residency**
1940	2.87	0.76
1950	2.48	0.91
1960	2.19	—
1970	1.74	1.13
1980	1.69	1.18
1990	1.60	1.17
2000	1.59	1.15

Notes: Authors' calculations, 1940 to 2000 PUMS. The southern residency index is the ratio of fraction black men aged 25 to 55 inclusive to white men of the same age. The MSA residency index is, for similar aged men, the ratio of black men to white men residing in an MSA. The 1970 southern residency index is created using the 1 percent state sample rather than the metro sample used in the rest of this paper because region of residency is not uniquely identified for MSAs that straddle the border of the southern region.

Table 4. Black-White Gaps in Log Weekly Wages, by City

Cities	**1940**	**1950**	**1970**	**1980**	**1990**	**2000**
Southern						
Houston	-0.834	-0.367	-0.425	-0.270	-0.297	-0.329
Memphis	-0.742	-0.454	-0.461	-0.319	-0.343	-0.248
Atlanta	-0.687	-0.472	-0.482	-0.288	-0.293	-0.253
New Orleans	-0.669	-0.612	-0.460	-0.317	-0.298	-0.286
Eastern						
Washington	-0.569	-0.374	-0.333	-0.225	-0.214	-0.187
New York	-0.478	-0.344	-0.282	-0.276	-0.268	-0.248
Philadelphia	-0.462	-0.394	-0.237	-0.259	-0.262	-0.266
Baltimore	-0.444	-0.333	-0.341	-0.254	-0.280	-0.241
Midwestern						
St. Louis	-0.515	-0.456	-0.298	-0.261	-0.309	-0.261
Cleveland	-0.507	-0.303	-0.280	-0.205	-0.221	-0.265
Chicago	-0.452	-0.331	-0.283	-0.282	-0.303	-0.306
Detroit	-0.376	-0.237	-0.208	-0.188	-0.206	-0.244
Western						
Los Angeles	-0.486	-0.283	-0.302	-0.264	-0.216	-0.229
San Francisco	-0.345	-0.255	-0.254	-0.236	-0.202	-0.222

Source: Authors' calculations, 1940 to 2000 I-PUMS.

TABLE 5. Log Wage Gaps for Blacks and Hispanics (Relative to Non-Hispanic Whites), 1990 Data from the NLSY 1979 Cohort

Men	(1)	(2)	(3)	(4)
Age	Yes	Yes	Yes	Yes
AFQT	No	No	Yes	Yes
Location Fixed Effects	No	Yes	No	Yes
Black	-0.246	-0.283	-0.062	-0.128
	(0.021)	(0.024)	(0.022)	(0.023)
Hispanic	-0.110	-0.163	0.017	-0.039
	(0.025)	(0.029)	(0.025)	(0.028)
N	4,206	4,206	4,206	4,206
Women	(1)	(2)	(3)	(4)
Age	Yes	Yes	Yes	Yes
AFQT	No	No	Yes	Yes
Location Fixed Effects	No	Yes	No	Yes
Black	-0.150	-0.177	0.086	0.028
	(0.021)	(0.025)	(0.021)	(0.026)
Hispanic	-0.027	-0.110	0.153	0.059
	(0.025)	(0.031)	(0.024)	(0.031)
N	3,959	3,959	3,959	3,959

Source: Authors' calculations from the NLS-Y79. Reported are coefficients from a linear regression in which the dependent variable is the natural logarithm of the wage in the current or most recent job. The age variable is denoted in calendar months. The AFQT variable is the 1989 renormed score from the 1981 test. Heteroskedastic robust standard errors reported in parentheses.

Table 6. Log Wage Gaps for Blacks and Hispanics (Relative to Non-Hispanic Whites), 2006 Data from the NLSY 1997 Cohort

Men	(1)	(2)	(3)	(4)
Age	Yes	Yes	Yes	Yes
AFQT	No	No	Yes	Yes
Location Fixed Effects	No	Yes	No	Yes
Black	-0.162	-0.221	-0.108	-0.183
	(0.036)	(0.042)	(0.039)	(0.037)
Hispanic	-0.010	-0.086	0.029	-0.056
	(0.034)	(0.045)	(0.036)	(0.048)
N	2,496	2,496	2,496	2,496
Women	(1)	(2)	(3)	(4)
Age	Yes	Yes	Yes	Yes
AFQT	No	No	Yes	Yes
Location Fixed Effects	No	Yes	No	Yes
Black	-0.185	-0.144	-0.063	-0.048
	(0.028)	(0.034)	(0.032)	(0.041)
Hispanic	-0.086	-0.169	0.018	-0.084
	(0.034)	(0.040)	(0.034)	(0.042)
N	2,531	2,531	2,531	2,531

Source: Authors' calculations from the NLS-Y97. Reported are coefficients from a linear regression in which the dependent variable is the natural logarithm of the wage in the current or most recent job. The age variable is denoted in calendar months. The AFQT variable is the CHRR norming of the score from the 1997 test. Heteroskedastic robust standard errors reported in parentheses.

References

Acemoglu, Daron, 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics*, 111(3), 779-804.

Angrist, Joshua and Alan Krueger, 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106(4), 979-1014.

Barron, John, Mark Berger, and Dan Black, 1997. "How Well Do We Measure Training?" *Journal of Labor Economics July*, 15(3), 507-528.

Becker, Gary S., 1964. Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. New York: National Bureau of Economic Research.

———, 1967. Human Capital and the Personal Distribution of Income: An Analytic Approach. Ann Arbor, Michigan: University of Michigan Press.

Beeson, Patricia E., 1991. "Amenities and Regional Differences in Returns to Worker Characteristics," *Journal of Urban Economics*, 30, 224-241.

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.

Black, Dan, Seth Sanders, and Lowell Taylor, 2003. "Measurement of Higher Education in the Census and CPS," *Journal of the American Statistical Association*, 98.

Bollinger, Christopher R., 2003. "Measurement Error in Human Capital and the Black-White Wage Gap," *Review of Economics and Statistics,* 85(3), 578-85.

Bound, John and George Johnson. 1992. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *American Economic Review*, 82(3), 371-92.

Card, David, 1998. "The Causal Effect of Education on Earnings" in Orely C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, 3A. Amsterdam: Elsevier Science, 1801-63.

———, 2001. "Estimating The Returns to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127-60.

Card, David and Alan Kruger, 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment," *Quarterly Journal of Economics* 107, 151-200.

Chandra, Amitabh, 2000. Labor Market Dropouts and the Racial Wage Gap, 1940-1990. Unpublished dissertation, University of Kentucky

Chen, Yong and Stuart Rosenthal, 2005. "Local Amenities and Life Cycle Migration: Do People Move for Jobs or Fun?" Unpublished paper, Syracuse University.

Chiswick, B (1974), *Income Inequality: Regional Analysis within a Human Capital Framework*, New York: Columbia University Press.

Council of Economic Advisers (1997), *Economic Report of the President*, Washington D.C.

Dahl, Gordon, 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica* 70, 2367-420.

Gabriel, Stuart and Stuart Rosenthal, 2004. "Quality of the Business Environment Versus Quality of Life: Do Firms and Households Like the Same Cities?" *Review of Economics and Statistics*, 86(1), 438-444.

Glaeser, Edward L. and Joseph Gyourko, 2005. "Urban Decline and Durable Housing," *Journal of Political Economy*, 133(2), 345-375.

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

Gyourko, Joseph, Matthew Kahn, and Joseph Tracy, 1999. "Quality of Life and Environmental Comparisons" in Edwin S. Mills and Paul Cheshire, eds., *Handbook of Regional and Urban Economics.*

Harmon, Oskar, 1988. "The Income Elasticities of Demand for Single-Family Owner-Occupied Housing: An Empirical Reconciliation," *Journal of Urban Economics*, 24, 173-185.

Haurin, Donald R., 1980. "The Regional Distribution of Population, Migration, and Climate," *Quarterly Journal of Economics*, 95(2), 293-808.

Hausman, J. A., Newey, W. K., and J. L. Powell, 1995. "Nonlinear Errors in Variables Estimation of Some Engel Curves," *Journal Of Econometrics*, 65, 205-233.

Heckman, James, Lochner, Lance, and Petra Todd, 2003. "Fifty Years of Mincer Earnings Regressions," Working Paper 9732, National Bureau of Economic Research.

Horowitz, Joel L., 1998. *Semiparametric Methods in Econometrics.* Berlin: Springer-Verlag.

Horrace, William C. and Peter Schmidt, 2000. "Multiple Comparisons with the Best, with Economic Applications," *Journal of Applied Econometrics*, 15(1), 1-16.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, 1993. "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101(3), 410-442.

Katz, Lawrence F. and Kevin M. Murphy, 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107(1), 35-78.

Mincer, Jacob, 1974. *Schooling, Experience, and Earning.* New York: National Bureau of Economic Research.

Murphy, Kevin M. and Finis Welch, 1992. "The Structure of Wages," *Quarterly Journal of Economics*, 107(1), 285-326.

Neal, Derek A. and William R. Johnson, 1996. "The Role of Premarket Factors in Black-White Differences," *Journal of Political Economy*, 104(5), 869-95.

Neal, Derek, 2004. "The Measured Black-White Wage Gap among Women Is Too Small," *Journal of Political Economy*, 104(5), 869-895.

Olsen, Edgar, 1987. "The Demand and Supply of Housing Service: A Critical Survey of the Empirical Literature" in Edwin S. Mills, ed.,*Handbook of Regional and Urban Economics*, Elsevier Science.

0'Neill, June, 1990. "The Role of Human Capital in Earnings Differences Between Black and White Men," *Journal of Economic Perspectives*, 4(4), 25-45.

Roback, Jennifer, 1982. "Wages, Rents, and the Quality of Life," *The Journal of Political Economy*, 90(6), 1257-1278.

Rosen, Harvey S., 1985. "Housing Subsidies: Effects on Housing Decisions, Efficiency, and Equity," in *Handbook of Public Economics*, North-Holland, 375-420.

Smith, James P. and Finis Welch, 1989. "Glack Economic Progress After Myrdal," *Journal of Economic Literature,* 27(2), 519-564.

Western, Bruce. 2006. Punishment and Inequality in America. New York: Russell Sage Foundation.

Willis, Robert J., 1986. "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," in Orely C. Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics* 1. Amsterdam: Elsevier Science, 525-602.