5-2012

Measurement Error and the Black-White Wage Differential

Jared Huff
Clemson University, jaredmhuff@gmail.com

Measurement Error and the Black-White Wage Differential

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Economics

by
Jared Michael Huff
May 2012

Accepted by:
Dr. Thomas A. Mroz, Committee Chair
Dr. Michael T. Maloney
Dr. Charles J. Thomas
Dr. Paul W. Wilson

Abstract

This dissertation encompasses two papers. The first paper examines the impact of measurement errors in potential experience and reported education on estimates of the Black-White wage differential. I show that measurement error is not mean zero, is distributed differently for Black and White males and, for experience, is correlated with the value of the variable measured with error. Possible conditional distributions of the true values of education and experience, given reported values, are estimated using auxiliary data. This paper introduces a Maximum Likelihood method to deal with these errors, and evaluates this method as well as other methods currently used in the gender wage differential literature. The use of the Maximum Likelihood estimation method along with traditional Multi-Sample Two-Stage Least Squares reveals that a significant portion of the estimated Black-White wage differential in a classic Mincer-style regression is due to measurement error in reported educational attainment and (especially) potential experience. Use of predicted or probabilistic measures in lieu of reported education and potential experience reduce the estimated racial wage gap in the 2000 Census from 34 percent to less than 20 percent.

In the second paper, I examine how the introduction of competition from Southwest Airlines affects airfares in a variety of market structures. While the consensus of the airline literature is that entry by Southwest results in substantially reduced fares on the entered and nearby routes (known as the Southwest Effect), little attention has been paid to the differing effects across routes. This paper fills this hole in the literature in two ways. First, I use difference-in-difference estimates to determine the causal effect of Southwest entry on fares using a natural experiment (the repeal of the Wright Amendment) that allowed for competition from Southwest on routes where such competition was previously illegal. I show that, consistent with the literature, the average effect on fares of competition from Southwest is substantial. However, the per-route effect varies substantially, from a fall in fares of roughly 40% to a slight increase in fares. A fixed-effects regression centered around

the repeal uncovers some of the factors behind this difference. Specifically, the presence of existing low-cost carriers and the presence of the ticketing airline in the origin and destination airports are the most important factors behind the "Southwest Effect," while the route Herfindahl and ticketing airline's share of passengers on the route matter little. These results illustrate the hazards of using a small-scale case-study approach to estimating the fare effects of entry by Southwest, as well as a need for a deeper understanding behind the mechanics of the Southwest Effect.

Dedication

I dedicate this dissertation to my wife, Nancy Huff, my parents, Jerry and Felicia Huff, and my Savior, Jesus Christ. Without my wife, this dissertation and this author would be substantially lessened. Without my parents, I would not have the drive and curiosity requisite to start graduate school, much less finish it. Finally, without my Savior, I would find that "*Everything is futile*" (Ecclesiastes 8:18b, HCSB).

Acknowledgments

This dissertation has been greatly improved by the contributions of many people. First and foremost, I would like to thank my advisor, Tom Mroz, whose input and advise have improved this dissertation substantially. Without his time and dedication, much of this dissertation would not have been possible. I would also like to thank the other members of my committee. Chuck Thomas read through numerous revisions of this dissertation, especially the second chapter. His comments on these drafts were extraordinarily helpful in the construction of the final product. Mike Maloney also commented on several drafts of this dissertation, and his comments always led to both improved economic analysis and improved clarity. Finally, Paul Wilson was helped not just through comments on this dissertation, but also though his graduate Econometrics classes which laid the foundation for the knowledge I needed to complete this dissertation.

Also, I would like to thank Nancy Huff for her many hours of work proofreading and editing this dissertation. Finally, I would like to thank all of the participants of the Labor and Industrial Organization Workshops at Clemson University. The numerous comments and moral support that I received from these workshops were invaluable.

Any remaining errors are my own.

Table of Contents

List of Tables

List of Figures

Chapter 1

Measurement Error and the Black-White Wage Differential

1.1 Introduction

This paper examines the consequences of using potential experience and reported education on estimates of the racial wage differential[1]. It builds on three decades' worth of work that investigates the impact of errors in potential experience on the gender wage gap (e.g., Garvey and Reimers, 1980; Blau and Kahn, 2011). A handful of studies have documented measurement error present in reported education, particularly in the Census (e.g., Black et al., 2003), or in potential experience measures in the NLSY for White and minority men (e.g., Antecol and Bedard, 2004). This paper adds to the existing literature by examining in detail the impact of these errors on the racial wage gap using Census and auxiliary data. The analysis is aided by the use of an estimation technique that formally accounts for data misreporting, through the use of auxiliary data, by estimating the mapping between accurate variables and those measured with error and then building this mapping into the likelihood function.

The goal of this paper, and the goal of most papers on either the gender or racial wage differential (see Altonji and Blank, 1999 for a broad survey of this literature) are focused on the conditional wage differential. There are actually two forms of differentials, as discussed in length

[1] Potential experience is defined as age minus years of education minus 6.

by Cain (1986) – the conditional wage differential and the unconditional wage differential. The unconditional (or long-run) racial wage differential simply measures the difference in average wages between all Black and White males in the economy. Since, in the long run, if there are no inherent differences between the races we would expect to see no differences in wages, even differences due to differences in factors such as educational attainment and experience, as these should converge across time. If there is any long-run difference in wages between the groups, that difference must be due to outside factors (such as discrimination).

Focusing on the unconditional wage differential, of course, presents two problems. First, the estimation of the unconditional wage differential is relatively straightforward – it's just the difference in mean wages, and can be obtained by summary statistics - and so a literature is unlikely to form around the estimation thereof. Indeed, the unconditional wage differential for 22-to-44 year old Black and White males can be inferred from the summary statistics table of this paper (Table 1.2), and is about 40%. The second and (in this author's opinion) considerably more important problem with focusing on the unconditional wage differential is the inability of the unconditional wage differential to guide policy whose aim is the reduction of the differential. If we want to craft policy that aims to cure either the wage differential itself or the underlying differences (such as differences in educational attainment and experience accumulation) that the wage differential is a symptom of, the unconditional wage differential does not guide us to the identification of these underlying differences.

We must turn, then, to the conditional wage differential, which is the form of the wage differential that is estimated in this and most of the other papers on this project. The conditional wage differential is the difference in wages estimated after differences in "relevant" factors have been accounted for (where, naturally, the factors that are relevant differ greatly across studies). Generally, when factors like educational attainment and experience are used in the regression, the estimated conditional wage differential falls is below, and often substantially below, the unconditional wage differential. This is because, as numerous studies have shown (again, see Altonji and Blank, 1999), white males tend to be more educated and have higher labor force attachment. Accounting for the differences in education and experience, then, is able to, for lack of a better word, "explain" some of the wage differential, and we see a lower conditional wage differential. The lower conditional wage differential, however, does not mean that there is a lower difference in wages (or less discrimination faced by Blacks or females) – we know that from the unconditional wage differential. What the lower conditional wage differential instead says is that we have discovered some of the reason why Black males (for example) get paid less. In a discrimination framework, even if, after accounting for factors such as education and experience, we estimate a conditional wage differential of zero, we have not shown that no discrimination exists. Instead, what we have shown is that no discrimination of a specific type exists – in this example, there would be no discrimination in which two equally skilled workers of different races get paid different amounts, on average. Instead, what we have shown is that the discrimination takes place in forms other than pay directly, such as reduced access to schooling compared to Whites, or through employers having a racial bias on decisions such as layoffs and promotions.

So, what we can see is that estimating a conditional wage differential leads to very specific policy prescriptions. If accounting for experience differences generates a conditional wage differential that is half of the unconditional wage differential, then experience differences is a huge factor in the equality of wages between the races, and is thus a prime target for policy. If, however, accounting for experience leads to little reduction in the racial wage differential, then policy whose attempt is to minimize the racial wage differential or wage discrimination is probably best directed elsewhere.

This ability to better correct policy is what lies at the heart of this paper. Past researchers have used levels of education and experience in estimates of the Black-White wage differential[2]. When precise measures of education and experience are not available, potential experience and reported education are used in their place. I show that doing so is a mistake. The use of reported education and (especially) potential experience to estimate the Black-White wage differential results in an estimated differential that is substantially inflated compared to that obtained using better measures of education and experience. This inflated differential is, in turn, going to have consequences on the importance of addressing differences in education and experience in our differential-reducing policy prescription. Specifically, the use of reported education and potential experience makes differences in education and experience seem like considerably less important factors than they actually are.

While some researchers object to the use of experience in analyses of the racial wage differential (see Cain, 1986 for a discussion of the short-run and long-run wage differential, as well as Neal and Johnson, 1996, and Bollinger, 2003), studies continue to find estimates of the racial

[2]to avoid redundancy, I will use the term wage differential from this point forward to indicate the conditional wage differential

wage gap conditional on experience important (e.g., Altonji and Blank, 1999; Antecol and Bedard, 2004). Given the amount of research that has been done on such errors for the gender wage gap, the relative dearth of interest concerning the error introduced by using potential experience on the racial wage gap is surprising. Antecol and Bedard (2004) present an explanation for this lack of interest: "Although it has long been agreed that potential experience is a poor proxy for actual experience for women, many view it as an acceptable approximation for men." They proceed to show that this is a poor assumption for individuals in the NLSY. This paper has similar findings: Black and White males accumulate experience at different rates, especially for non-college graduates, and regardless of the auxiliary dataset used. The difference in experience accumulation rates has an effect not just on the estimated Black-White wage differential, but also on estimates of the correlation between wages and both education and experience.

Current methods in the gender wage differential literature involve the use of auxiliary data to predict actual experience using potential experience. I introduce a method that also focuses on the use of auxiliary data. However, unlike previous methods which use an explicit first stage to estimate a prediction equation for the mismeasured variables, I use the auxiliary data to estimate the distribution of true values of education and experience given their observed values (equivalent to calculating the distribution of the measurement errors). Once these distributions have been obtained, they can be directly accounted for in the likelihood function.

The use of reported education and (especially) potential experience significantly biases estimates of the Black-White wage differential: the differential estimated in the Census falls from 34.5 percent when using potential experience and reported education to no more than 20 percent when the more accurate approaches are used. This has several implications. First, using potential experience as a proxy for accumulated experience can impact empirical results in a statistically and economically significant way. Reported levels of educational attainment should also be viewed with suspicion. Further, the results of this paper suggest that influential classic studies (e.g., Smith and Welch, 1989) which estimate the wage gap using these measures in Census data likely overstate the racial wage differential, and possibly miss important reasons for the wage convergence between Black and White males between 1950 and 1980. Finally, these results illustrate the usefulness of an estimation technique that allows for a general distribution of the measurement error. The ability to bring more accurate measures of education and experience into the Census data allows for the use of Census data in a wider range of analyses.

1.2 Empirical Methods

To estimate the Black-White wage differential in the Census, I use a basic Mincer regression model:

$$ln(\text{income}_i) = \alpha_0 + \alpha_1 \text{EXP}_i + \alpha_2 \text{EXP}_i^2 + \alpha_3 \text{EDUC}_i + \alpha_4 \text{Black}_i + e_i, \tag{1.1}$$

where income is measured as total annual income, EXP and EXP^2 are an individual's accumulated experience level and its square, respectively, and $EDUC$ measures the educational attainment level of the individual. As I show later in this paper, the experience and education terms are measured with error (through use of potential experience, denoted PE, and reported education), which leads to bias in the estimated coefficients for education and experience. Further, these errors are distributed differently for Whites and Blacks, leading to a biased estimator for the racial difference in wages. As the measurement errors are correlated with the value of the variables being measured with error, these errors cannot be corrected by using standard IV methods. Given the existence of auxiliary data that give accurate measures of education and experience, three estimation procedures are considered - a Multi-Sample Two-Stage Least Squares (2SLS) method, a regression technique introduced in Regan and Oaxaca (2009), and a Maximum Likelihood method that relies on knowledge of the distribution of the errors in the independent variables. The former two have been previously used in the literature, and the third is novel to this paper.

Multi-Sample 2SLS has been used in a variety of applications (see Inoue and Solon, 2010 for a limited survey), including wage differentials (e.g., Filer, 1993). When used in wage differential analyses, the first stage uses auxiliary data with measures of experience and variables that are common to the database with out a reliable experience measure to create a prediction function (of experience) which can then be used in the primary dataset to generate predicted values of experience. In my analysis, three first-stage regressions are necessary – one for each of the three problematic explanatory variables:

$$EXP_i = \alpha_0 + \alpha_1 PE_i + \alpha_2 PE_i^2 + \alpha_3 \text{reported education}_i + \alpha_4 Black_i + e_i \tag{1.2}$$

$$EXP_i^2 = \beta_0 + \beta_1 PE_i + \beta_2 PE_i^2 + \beta_3 \text{reported education}_i + \alpha_4 Black_i + u_i \tag{1.3}$$

$$EDUC_i = \gamma_0 + \gamma_1 PE_i + \gamma_2 PE_i^2 + \gamma_3 \text{reported education}_i + \alpha_4 Black_i + \epsilon_i. \tag{1.4}$$

Avoiding the Forbidden Regression (discussed in Wooldridge, 2002, among other places) in the second stage requires separately predicting experience and its square. The first-stage coefficients obtained from these regressions are then used to predict experience, the square of experience, and education in the Census, denoted as $\widehat{EXP}$, $\widehat{EXP^2}$, and $\widehat{EDUC}$, respectively, from reported education and potential experience.

Note here that one of Equations 1.2 and 1.3 *must* be misspecified – if $Y_i = X_i\beta + e_i$, where e_i is normally distributed, then $Y_i^2 = (X_i\beta + e_i)^2 \neq X_i\gamma + u_i$, where u_i is normally distributed. Crucially, however, 2SLS does not depend on a correct first-stage specification. Instead, 2SLS simply requires orthogonality between the instruments and the error term in the second stage.

The second stage regression simply consists of regressing log income on $\widehat{EXP}$, $\widehat{EXP^2}$, and $\widehat{EDUC}$, as well as the racial indicator variable. The resulting coefficients give an asymptotically unbiased estimate of the true coefficients. Because error is introduced through the first stage, the standard errors of this estimator need to be corrected. The method by which I estimate the standard errors is discussed in the next subsection. The properties of two-step estimators in general are discussed in detail in Murphy and Topel (1985), with Two-Sample 2SLS specifically discussed in Inoue and Solon (2010).

The second method is proposed by Regan and Oaxaca (2009), who express concerns about negative predicted values of experience in linear first-stage regressions such as Equations 1.2-1.4 above. To bound predicted values of experience away from zero, they use a slightly different procedure. Instead of two first-stage regressions for experience, they have only one: ln(experience) regressed on age and the remaining second-stage independent variables. Since my analysis is concerned with measurement error in reports of educational attainment as well as in potential experience, the two first-stage regressions of this method for this paper are then:

$$ln(EXP_i) = \alpha_0 + \alpha_1 age_i + \alpha_2 \text{reported education}_i + \alpha_3 Black_i + e_i \tag{1.5}$$

$$EDUC_i = \gamma_0 + \gamma_1 age_i + \gamma_2 \text{reported education}_i + \alpha_3 Black_i + \epsilon_i. \tag{1.6}$$

Predicted values for the log of experience are obtained both in the Census and in the auxiliary data. These predicted log-experience values can then be transformed into a predicted experience measure by the transformation $e^{(\widehat{ln(EXP)})}$. Since the mean transformed predicted values in the auxiliary data will not equal the sample mean of actual experience in the auxiliary data, a multiplicative scalar, ζ, is calculated that equates these two means. This scalar can then be applied to the transformed predicted values in the Census data, such that predicted experience in the Census is equal to $\zeta e^{(\widehat{ln(EXP)})}$. From this, the predicted value for the square of experience can then be calculated by squaring predicted experience for each observation.

The problem with the Regan-Oaxaca model is the following. Consider a simple model where the truth is

$$y_i = X_i\beta + \epsilon_i.$$

X_i is not observed, but is a function of an observed variable Z_i that only affects y_i through its effect on X_i. Assuming that the first-stage proposed by Regan-Oaxaca is correct, it must be the case that $ln(X_i) = Z_i\gamma + v_i$, and thus $X_i = e^{Z_i\gamma}e^{v_i}$. We get estimates $ln(X_i) = Z_i\hat{\gamma} + \hat{v}_i$. The true model for y_i, then, is

$$y_i = \alpha + e^{Z_i\gamma}e^{v_i}\beta + \epsilon_i = \alpha + e^{Z_i\hat{\gamma}}e^{\hat{v}_i}\beta + \epsilon_i.$$

Unfortunately, we don't observe γ or v_i in the first stage. Instead, we use the estimate $\hat{\gamma}$ to predict $ln(X_i)$ in the second stage. If the first and second stage are in different samples, the second stage $\hat{v}_i$ are unobservable, and so we need to make a correction when turning $\widehat{ln(X_i)}$ into $\hat{X}_i$. We do this using the adjustment factor mentioned earlier. This adjustment factor, $\zeta = \overline{X}/\overline{\hat{X}}$, can be be written as

$$\zeta = E[e^{Z_i\gamma}e^{v_i}]/E[e^{Z_i\hat{\gamma}}].$$

Since $Z_i\widehat{\gamma} = Z_i\gamma + v_i - \hat{v}_i$, the adjustment factor can be simplified to

$$\zeta = E[e^{Z_i\gamma}e^{v_i}]/E[e^{Z_i\gamma}e^{v_i}/e^{\hat{v}_i}] = E[e^{\hat{v}_i}].$$

So, our second stage regression is

$$y_i = \alpha + [e^{Z_i\widehat{\gamma}}E[e^{\hat{v}_i}]]\tilde{\beta} + \epsilon_i.$$

Modifying the true model means to reflect what we are actually estimating reveals the following:

$$y_i = \alpha + e^{Z_i\widehat{\gamma}}E[e^{\hat{v}_i}]\beta + \eta_i,$$

where

$$\eta_i = [\beta(e^{Z_i\hat{\gamma}}e^{\hat{v}_i} - e^{Z_i\hat{\gamma}}E[e^{\hat{v}_i}]) + \epsilon_i].$$

So, when we get the estimate

$$y_i = \hat{\alpha} + e^{Z_i\widehat{\gamma}}E[e^{\hat{v}_i}]\hat{\beta} + \hat{\eta}_i,$$

it is the case that $E[\hat{\beta}] = \beta$ only when $cov(e^{Z_i\hat{\gamma}}e^{\hat{v}_i} - e^{Z_i\hat{\gamma}}E[e^{\hat{v}_i}]) = 0$. In general, heteroskedastic errors will invalidate these assumptions.

If the true model for y_i also contains a squared experience term,

$$y_i = \alpha + X_i\beta_1 + X_i^2\beta_2 + \epsilon_i = \alpha + e^{Z_i\hat{\gamma}}e^{\hat{v}_i}\beta_1 + e^{2Z_i\hat{\gamma}}e^{2\hat{v}_i}\beta_2 + \epsilon_i,$$

then additional problems arise in the estimation of $\hat{\beta}_2$. Since $\widehat{X_i^2} = (\widehat{X_i})^2$ by construction, instead of

$$X_i^2 = e^{(Z_i\gamma)^2}e^{v_i^2} = e^{2Z_i\gamma}e^{2v_i},$$

we use

$$\widehat{X_i^2} = e^{2Z_i\hat{\gamma}}E[e^{\hat{v_i}}]^2.$$

Letting

$$y_i = \alpha + e^{Z_i\hat{\gamma}}E[e^{\hat{v}_i}]\hat{\beta}_1 + e^{2Z_i\hat{\gamma}}E[e^{\hat{v}_i}]^2\hat{\beta}_2 + \epsilon_i.$$

be the equation that we estimate and focusing just on $\hat{\beta}_2$, $E[\beta_2] - \beta_2 = 0$ only when

$$E[e^{Z_i\hat{\gamma}}E[e^{\hat{v}_i}]^2 - e^{Z_i\hat{\gamma}}e^{2\hat{v}_i}] = 0.$$

Note, however, that even the terms involving v_i are no longer equal in expectation (even assuming independence of $e^{Z_i\hat{\gamma}}$ and $e^{\hat{v}_i}$). For example, when ϕ is a normally distributed random variable, e^{ϕ} is log normal and has an expected value equal to $e^{(\frac{1}{2}\sigma_\phi^2)}$, where σ_ϕ^2 is the variance of ϕ. The expected value of $e^{2\hat{v}_i}$, then, is $e^{\frac{1}{2}4\sigma_v^2} = e^{2\sigma_v^2}$. The expected value of $E[e^{\hat{v}_i}]^2$, on the other hand, is

$(e^{\frac{1}{2}\sigma_v^2})^2 = e^{\sigma_v^2}$. As $e^{2\sigma_v^2} \neq e^{\sigma_v^2}$, $E[\hat{\beta}_2] \neq \beta_2$.

It is worth noting that, while the estimated coefficients are unlikely to be correct using this method, the predicted $\hat{y}_i$ from the second stage will be the same as when X_i is directly observable. However, if we are interested in the experience-earnings profile (as is the case with wage differentials, since the shape of the experience-earnings profile will affect the estimated differential), this method is likely to lead to biased estimates.

An additional way to incorporate the information gained from use of auxiliary data into the Census is the through use of a different empirical method unique to this paper: a Maximum Likelihood technique that builds the uncertainty directly into the estimation. A typical OLS-equivalent log-likelihood function using observed values and assuming normally distributed errors can be written as

$$llf_i = ln(\phi(\frac{y_i - \mathbf{X}_i\beta}{\sigma})) - ln(\sigma),$$

where ϕ is the standard normal density function. In the presence of measurement error in $\mathbf{X}$, the estimated coefficient, $\hat{\beta}_{MLE} = \hat{\beta}_{OLS}$, is biased.

If we know the distribution of measurement error in $\mathbf{X}$, however, a correct conditional log-likelihood can be written explicitly. Intuitively, suppose that observation i has, for $\mathbf{X}$, an observed value of 2. Conditional on this observed value there is a probability of 0.3 that the true value is 2 and a probability of 0.7 that the true value is 3. Assuming a normal error term and again letting y be the dependent variable, the likelihood function of observation i is thus

$$lf_i = 0.3 * \frac{1}{\sigma}\phi(\frac{y_i - 2\beta}{\sigma}) + 0.7 * \frac{1}{\sigma}\phi(\frac{y_i - 3\beta}{\sigma}).$$

This can be rewritten as

$$\begin{aligned} lf_i &= \sum_{m=2}^{3} pr(X_i = m | X_{i,obs} = 2)\frac{1}{\sigma}\phi(\frac{y_i - m\beta}{\sigma}) \\ &= \sum_{m=2}^{3} \pi_{im}\frac{1}{\sigma}\phi(\frac{y_i - m\beta}{\sigma}), \end{aligned}$$

where m represents the possible true values of $\mathbf{X}$. When the values of $\mathbf{X}$ are discrete, a general form can be written as (after taking logs)

$$llf_i = ln(\sum_j \pi(\mathbf{X}_{i,obs})_{ij}\phi(\frac{y_i - \mathbf{X}_{ij}\beta}{\sigma})) - ln(\sigma), \tag{1.7}$$

where $\mathbf{X}_{i,obs}$ denotes the observed values for $\mathbf{X}$ for observation i. The $\pi(\mathbf{X}_{i,obs})_{ij}$ above represent the probability of some set of values occurring for variables $\mathbf{X}$ given the observed values for $\mathbf{X}$, where the j term indexes each possible set of values. The $\pi(\mathbf{X}_{i,obs})_{ij}$ are obtained from an examination of the auxiliary datasets. In the estimation conducted later in this paper, the variables in $\mathbf{X}$ are the same as in the previous two methods: education, experience, experience squared, and a racial indicator variable. The $\pi(\mathbf{X}_{i,obs})_{ij}$, then, represent the probability that an individual has various true-education-experience-experience-squared values given the observed values of education, potential experience and its square, and his race, which can then be integrated (or, if discrete, summed) over. An advantage of this method is that it (like the Regan & Oaxaca method) eliminates "predictions" of values outside of feasible ranges, which previous researchers have found appealing. Further, if the measurement errors are known with certainty, then the probabilities can be calculated with certainty and the MLE estimates are efficient (as it asymptotically achieves the Cramer-Rao Lower Bound). A disadvantage is the necessity of observing (or estimating) the joint distribution of the error terms, or, equivalently, the probabilistic relationship between observed and true values. While this may be impractical in many situations, it is relatively straightforward here due to the same auxiliary data that makes possible estimation of the Multi-Sample 2SLS and Regan & Oaxaca approaches.

1.2.1 Estimating the Conditional Probabilities for the MLE

In general, the $\pi(\mathbf{X}_{i,obs})_{ij}$ will be unknown, and thus need to be estimated. In both the Monte Carlo simulation and the empirical analyses conducted later, the probabilities are estimated in the following manner. First, auxiliary samples containing both accurate measures of education and experience as well as measures similar to those in the primary databases are obtained (the sample for education and the sample for experience need not be the same).[3] After the "true" experience values have been calculated in the auxiliary data, I smooth the relationship between the true and

[3]The manner in which these auxiliary databases are generated are discussed in the appropriate sections of this paper.

reported values of experience using a hazard model described below. This smoothing process has the advantage of using information from within-race groups with a large number of observations in the assignment of probabilities for those groups with fewer observations. This helps eliminate observed phenomena that seem unlikely, such as groups of individuals having a zero probability of 50-57 years of experience, but a positive probability of 58 years of experience. Further, an examination of the maximized log-likelihood function values obtained using smoothed and unsmoothed probabilities show that the smoothed probabilities lead to a better "fit" of the data, as the maximized log-likelihood function values are closer to zero.

The probabilities are smoothed in a process similar to Gilleskie and Mroz (2004). Each individual is observed to have some true level of experience between 0 and 59 full-time equivalent years. That individual is treated as being at risk of "failure," where failure is defined as achieving a specific level of true experience (i.e., his observed value). After failure, the individual leaves the sample. The hazard function (that is, the probability of an individual having some level of experience given that he does not have a lower level of experience) depends on the log age, race, and education level of the individual, as well as the level of experience under consideration. To allow for a flexible estimated hazard function, a total of 29 independent variables are used, consisting of the variables mentioned above and interactions and exponents thereof.[4]

Once the hazard function is obtained, I calculate the survival and cumulative survival functions. The predicted probability that an individual has a specific level of experience is then the probability that he does not have a lower level of experience (obtained from the estimated cumulative survival function) times the conditional probability that he has that level of experience (obtained from the estimated hazard function).

For both databases that I use, this method was implemented as follows. First, each observation was duplicated 59 times. Each observation in the dataset, then, represents a candidate true experience value from ranging from 0 to 59, inclusive. An indicator variable is created that equals 0 if the true experience value was greater than this candidate value, 1 if the true experience and candidate experience values were the same, and was "missing" otherwise (to represent the observation

[4]The exact specification was chosen using AIC, BIC, and cross-validation, all of which indicated that the same model should be used. The 29 independent variables are, letting EXP and EDUC denote the true experience level under consideration and the education level of the individual, respectively, $log(age)$, $log(age)^2$, $log(age)^3$, $log(age)*EXP$, $log(age)*EXP^2$, $log(age)*EXP^3$, EXP, EXP^2, EXP^3, $Black$, $Black*log(age)$, $Black*log(age)^2$, $Black*log(age)^3$, $Black*EXP$, $Black*EXP^2$, $Black*EXP^3$, $Black*log(age)*EXP$, $Black*log(age)*EXP$, $Black*log(age)*EXP$, $EDUC$, $EDUC*log(age)$, $EDUC*log(age)^2$, $EDUC*EXP$, $EDUC*EXP^2$, $EDUC*EXP^3$, $EDUC*log(age)*EXP$, $EDUC*log(age)*EXP^2$, $EDUC*log(age)*EXP^3$, and a constant term.

having left the sample due to "failure"). A logit regression using all of the "outcomes" that are zero or one is then estimated, where this indicator variable is the dependent variable and is regressed on the 29 independent variables mentioned previously. I use these results to construct predicted probabilities, which equal the estimated conditional hazard rates. I construct the cumulative survival probabilities for all experience levels using the estimated hazard rates. This is equal to 1 for all people at zero years of experience (that is, no one could have achieved an experience level below zero years). The cumulative survival probability for all experience levels greater than zero is equal to the cumulative survival probability of the previous period times the conditional hazard rate of the previous period. So, for example, the cumulative survival probability for one year of experience is one (the cumulative survival probability of zero years of experience) times the probability, conditional on having no less than zero years of experience, that the person has exactly zero years of experience (the hazard rate of zero years of experience). The probability that a person has a level of experience between 0 years and 58 years, then, is the cumulative survival probability times the hazard rate. That is, the probability that a person has some level of experience x is equal to the probability that he has no fewer years of experience than x times the probability that, conditional on having no fewer years of experience THAN X, he has exactly x years of experience.

Formally, let $Haz(Exp_i|X_i)$ denote the hazard function, $Surv(Exp_i|X_i)$ denote the cumulative survival function, and $Prob(Exp_i|X_i)$ denote the probability density function. Estimating the logit model using the set of explanatory variables X and the explanatory variable Exp results in estimated coefficients $\hat{\alpha}$ for X_i and $\hat{\beta}$ for Exp_i[5]. The hazard function is calculated as

$$Haz(Exp_i|X_i) = e^{\hat{\alpha}X_i+\hat{\beta}Exp_i}/(1+e^{\hat{\alpha}X_i+\hat{\beta}Exp_i}). \tag{1.8}$$

The cumulative survival function, then, can be expressed as

$$\begin{aligned} Surv(Exp_i|X_i) &= 1 \text{ if } Exp_i = 0 \\ &= Surv(Exp_i - 1|X_i) - Haz(Exp_i - 1|X_i) \text{ if } Exp_i > 0. \end{aligned} \tag{1.9}$$

[5]See Appendix A for these estimated coefficients using the NLSY data

Finally, the probability density function can be expressed as

$$Prob(Exp_i|X_i) = Surv(Exp_i|X_i) * Haz(Exp_i|X_i). \tag{1.10}$$

To ensure that the probabilities sum to one, the probability that he has 59 years of experience is defined as one minus the sum of the other probabilities (which cannot exceed one by construction).

The first stage regression and select probabilities for the NLSY experience specification are shown in Appendix A. Included in the appendix are the results from the hazard regression and graphs illustrating the hazard function, cumulative survival function, and experience probabilities for two very different pairs of individuals — 40-year-old college graduates and 22-year-old high school dropouts. When examining the graphs, recall that the experience probability mapping is simply the product of the hazard and cumulative survival functions. The graphs for the two groups are quite different, as we would suspect: the mass of the experience distribution for the college graduates (both Black and White) is considerably farther to the right than is that of the high school dropouts, which is supported by the raw data from both the NLSY and PSID, which will be discussed in a later section.

The true education probabilities are calculated in a more straightforward manner. Since there are only four outcomes of interest, I use a simple multinomial logit model instead of a hazard function to estimate the conditional education probabilities. In this multinomial logit, the four possible true values are assumed to be a function of reported values for each race. Note that, with the multinomial logit, the predicted probabilities will simply equal the empirically observed conditional probabilities. However, use of the multinomial logit model will simplify the estimation of the standard error for the MLE approach as I discuss shortly.

1.2.2 Calculation of the Standard Errors of the Estimates

Estimation of the standard error of both the Maximum Likelihood Estimation and Multi-Sample 2SLS face the following problem. In each, I begin by minimizing some function to estimate some parameters of interest, i.e., $\min_{\beta_1} F_1(X_1\beta_1)$.[6] This estimation yields estimates $\hat{\beta_1}$. These estimates are then used in the estimation of *other* parameters of interest in a second-stage equation,

[6] Here, F is a generic function. We may be interested in minimizing the sum of squared residuals, for instance, or minimizing the negative likelihood function.

$\min_{\beta_2} F_2(X_1\beta_2|\hat{\beta_1})$.[7] Neglecting to account for the estimation error inherent in the $\hat{\beta}_1$ results in standard errors in the second stage that are too small. As such, these standard error need to be adjusted, and are done so using the method described below.

The standard errors of the MLE estimation are obtained using the process described in the appendix of Mroz (1987). The covariance matrix of a Maximum Likelihood estimate β can typically be calculated using a sandwich estimator,

$$cov(\beta) = H^{-1}[\sum_i(\frac{\partial F_i}{\partial \beta})(\frac{\partial F_i}{\partial \beta})']H^{-1},$$

where H is the empirically estimated Hessian matrix, F is the vector of maximized likelihood function values, and $\sum_i(\frac{\partial F_i}{\partial \beta})(\frac{\partial F_i}{\partial \beta})'$ is the matrix of outer partial derivatives. Since the estimation procedures in this paper have two stages that are done in different databases, the formula for the standard errors of the second-stage estimate of β_2 is (letting subscripts 1 and 2 denote first-stage and second-stage estimates, respectively):

$$cov(\beta_2) = H_2^{-1}[\sum_i(\frac{\partial F_{2i}}{\partial \beta_2})(\frac{\partial F_{2i}}{\partial \beta_2})']H_2^{-1} + H_2^{-1}[\sum_i \frac{\partial^2 F_{2i}}{\partial \beta_2 \partial \beta_1}]H_1^{-1}[\sum_i \frac{\partial^2 F_{2i}}{\partial \beta_2 \partial \beta_1}]'H_2^{-1}. \tag{1.11}$$

Here, $\frac{\partial^2 F_{2i}}{\partial \beta_{2i} \partial \beta_1}$ is the derivative of the first derivatives from the second-stage estimation procedure with respect to the first-stage coefficients used to get the conditional probabilities for observation i. The difference between the manner in which the standard errors for the Multi-Sample 2SLS and the Maximum Likelihood estimates are calculated lies in the estimation of this cross-partial derivative. Because the first stage of the Multi-Sample 2SLS is a straight-forward linear equation, the cross-derivative can be derived algebraically. The use of a hazard function (in the case of experience) and a multinomial logit procedure (in the case of education) makes deriving this term algebraically impractical. Instead, each of the cross-partial derivatives is approximated numerically in two parts. First, the sum of the observation-level first derivatives of each of the six (including σ) estimated second-stage coefficients are calculated using the likelihood function with the original education and experience probabilities. Next, new education and experience probabilities are calculated by deviating each of the 24 (in the case of education) or 29 (in the case of experience) first-stage coefficients by a small amount. The sum of the observation-level first derivatives are then calculated

[7] In Multi-Sample 2SLS, the $\hat{\beta_1}$ represent predicted values of education and experience. In the MLE, they represent predicted education and experience probabilities.

using the the original estimated second-stage coefficients and the "new" probabilities. Formally, letting $\tilde{\pi}(\hat{\beta}_1)$ be the probabilities calculated using estimated first stage coefficients $\hat{\beta}_1$ and $\tilde{\pi}(\hat{\beta}_{1,i} + h, \hat{\beta}_{1,-i})$ be the probabilities calculated when first stage coefficient for variable i, and only that coefficient, is deviated by h, the formula for the numerical cross-derivatives of observation i is

$$\frac{\partial^2 F_{2i}}{\partial\widehat{\beta_2}\partial\widehat{\beta_1}} = \left(\frac{\partial F_{2i}}{\partial\widehat{\beta_2}}\bigg|_{\tilde{\pi}(\widehat{\beta}_{1,i}+h,\widehat{\beta}_{1,-i})} - \frac{\partial F_{2i}}{\partial\widehat{\beta_2}}\bigg|_{\tilde{\pi}(\widehat{\beta}_1)} \right) /h. \tag{1.12}$$

The value of h is chosen by trying multiple candidates[8] until numerical stability in the estimated derivatives is observed.

1.2.3 Monte Carlo Evidence

To examine the performance of the estimation techniques, I design a simulation study in which education and experience are measured with error that is distributed differently for Blacks and Whites. The setup of the simulation is as follows. I want to obtain $\widehat{\beta}$ from the estimation of

$$y_i = \mathbf{x}_i\beta + \epsilon_i. \tag{1.13}$$

However, instead of directly observing $\mathbf{x}$, I only observe some vector of variables $\mathbf{z}$ which measure $\mathbf{x}$ with error. In the simulation, I obtain values of $\mathbf{z}$ from randomly drawn subsamples of the Census and use these to generate some "true" values of $\mathbf{x}$ according to the processes defined below. Once the vector $\mathbf{x}$ is generated, I generate values of y using Equation 1.13 above, where β is known. I then try to recover β through estimates, $\widehat{\beta}$, obtained by various estimation methods that use the "observed" values of $\mathbf{z}$ instead of the true values of $\mathbf{x}$. I repeat the process many times, and compare the average $\widehat{\beta}$ to the (known) true value of β.

I randomly draw 1,000 datasets with 10,000 observations each from the 2000 Census. These observations have values for race (Black or White), age (22-44 years), and reported education (10, 12, 14, or 16 years). I begin by generating "true" years of education based on the *reported* educational

[8] ranging from $max(0.01, 0.01 * |\beta_{1,i}|)$ to $max(0.00000001, 0.00000001 * |\beta_{1,i}|)$

attainment using the following:

$$White \begin{cases} \text{true education equals reported education}-4 \text{ years with 2.5\% probability} \\ \text{true education equals reported education}-2 \text{ years with 5\% probability} \\ \text{true education equals reported education with 88.75\% probability} \\ \text{true education equals reported education}+2 \text{ years with 2.5\% probability} \\ \text{true education equals reported education}+4 \text{ years with 1.25\% probability} \end{cases}$$

$$Black \begin{cases} \text{true education equals reported education}-4 \text{ years with 5\% probability} \\ \text{true education equals reported education}-2 \text{ years with 10\% probability} \\ \text{true education equals reported education with 77.5\% probability} \\ \text{true education equals reported education}+2 \text{ years with 5\% probability} \\ \text{true education equals reported education}+4 \text{ years with 2.5\% probability.} \end{cases}$$

True education equals reported education with the the probability necessary for the total probability to sum to one. True experience, denoted EXP, is then calculated as

$$EXP_i = \alpha_i \gamma_i (age_i - true\ education_i - 6), \tag{1.14}$$

where γ_i is randomly drawn from a uniform distribution ranging between 0 and 1 for each observation, and α_i differs across race and true education levels as

$$\alpha_{White} = \begin{cases} 1.7 & \text{if college graduate} \\ 1.6 & \text{if some college} \\ 1.5 & \text{if high school graduate} \\ 1.4 & \text{if high school dropout} \end{cases} \qquad \alpha_{Black} = \begin{cases} 1.6 & \text{if college graduate} \\ 1.45 & \text{if some college} \\ 1.35 & \text{if high school graduate} \\ 1.25 & \text{if high school dropout.} \end{cases}$$

The product $\alpha_i \gamma_i$ is used to mimic what is observed in the auxiliary data discussed in the next section: true experience values at each age-education-race grouping vary substantially, and there are generally individuals with experience values of zero as well as individuals who have accumulated more years of full-time equivalent work experience than potential experience would indicate. This latter effect seems to decline as education level falls. Finally, the square of experience is calculated as $EXP_i^2 = (EXP_i)^2$.

Once "true" $EDUC$, EXP, and EXP^2 have been generated, I use these variables, along with the racial indicator variable from the Census, to generate "true" $\log(wages)$:

$$\log(wages_i) = 7 + .15EXP_i - .0025EXP_i^2 + .15EDUC_i - .2Black_i + \epsilon_i. \tag{1.15}$$

The ϵ_i are randomly drawn from a standard normal distribution.

Several estimators are tested in this simulation. OLS with the "true" values is used as a benchmark, along with the known β. The first model tested is a naïve OLS estimation using potential experience variables and reported education in place of the accumulated experience variables and true education, respectively. The results of the naïve OLS estimation are shown in column 3 of Table 1.1. Note that the estimate of the conditional Black-White wage differential is dramatically larger (in absolute value) when reported education and potential experience are used instead of true education and accumulated experience. Further, the estimates of the correlation between earnings and both education and experience differ from the truth.

The next estimation method tested is the method proposed in Regan and Oaxaca (2009), with the results shown in column 4 of Table 1.1. As the Regan & Oaxaca method, Multi-Sample 2SLS, and MLE all use a auxiliary data, 1,000 additional databases are constructed in the manner discussed above. These additional databases serve as auxiliary databases in which the relationship between reported and "true" variables can be measured. While the Regan & Oaxaca method does much better than the naïve OLS estimation, the average point estimate of the primary coefficient of interest (β_{Black}) is still overestimated by over one percentage point. This method is outperformed by both the Multi-Sample 2SLS and the MLE method, shown in columns 5 and 6. The Multi-Sample 2SLS gives the most accurate average point estimate (-.201 compared to -.196), but the MLE has a slightly lower standard deviation (at the fourth decimal place).

In summary, Table 1.1 shows that Multi-Sample 2SLS and the MLE method more accurately estimate the true relationship than does a naïve OLS estimation using reported education and potential experience. Further, the unadjusted Multi-Sample 2SLS procedure more accurately estimates the correct values for the coefficients than does the Regan & Oaxaca procedure, even though Multi-Sample 2SLS often predicts negative values for experience. The MLE method does slightly worse at estimating the coefficients of interest than does Multi-Sample 2SLS (and performs better than Regan & Oaxaca for each of the coefficients of interest), but the standard deviation of the estimates is lower. These simulation results indicate that Multi-Sample 2SLS and the MLE method reduce much of the bias. As the Maximum Likelihood method also eliminates negative predicted values of experience and the Regan & Oaxaca method performs poorly in the simulation, the Regan & Oaxaca method is not used in the empirical analysis.

1.3 Data

I focus on young (aged 22-44) Black and White males in the 2000 Census. To make the groups as similar as possible in composition, I look only at individuals who do not identify themselves as Hispanic. Summary statistics of the relevant Census variables are shown in Table 1.2. White males tend to be slightly older than their Black counterparts, and have on average, very similar years of potential experience. When educational attainment is grouped into four categories (10 years for high school dropouts, 12 for high school graduates, 14 for individuals with some college, and 16 for college graduates), White males tend to have about a half-year more schooling than do Black males.

Several studies have either noted that Black males tend to have less experience on average than do White males, and to a greater degree than differences in potential experience might indicate (see Bratsberg and Terrell, 1998; Altonji and Blank, 1999; Antecol and Bedard, 2004), making use of the Census variables in Table 1.2 inappropriate for conditional wage gap estimations that want to hold account for experience differences. Further, Black et al. (2003) casts doubt on the accuracy of self-reported educational attainment in the Census by comparing select reported college graduates in the 1990 Census with the same individuals in an auxiliary database (the 1993 National Survey of College Graduates, or NSCG)[9]. They find considerable evidence of misreporting errors in reports of educational attainment that differ by gender and race.

For this analysis, I assume that the errors in education and potential experience are uncorrelated.[10] This assumption is made because of sample-size issues, especially in the education variables. If the data used to generate the sibling-reported education measure discussed below were used to simultaneously estimate the relationship between potential and accumulated experience, for example, I would either have to focus on a much narrower age range or omit age-specific differences

[9]The 1993 NSCG was a re-sampling of the pool of individuals who were reported as having obtained a college degree in the 1990 Census.

[10]Although I use education-corrected potential experience in the MLE. That is, when I give an individual some probability of having a high school diploma instead of their college degree, for example, I simultaneously increase their potential experience up by four years.

in education error. As an illustration, Table 1.8 shows that I observe 17 black college graduates who meet the criteria for inclusion in that specific sample. It is implausible to use these 17 observations to estimate the distribution of experience for 23 different ages — even with the first-stage smoothing of experience probabilities, the resulting standard errors would likely render the estimates useless in trying to make meaningful inference. For the sake of comparison, note that the "independent" NLSY experience measure has just over 2400 observations for black college graduates. Even with this "large" sample, the standard error on the estimated differential in both MLE and Multi-Sample 2SLS estimates that simultaneously account for education and experience errors is less than a quarter of the size of the estimated differential only once.

The following two sections approach each of these issues in turn. First, the degree to which potential experience fails to capture actual experience in the Census and other databases is analyzed. I then construct possible relationships between reported and true educational attainment, again making use of the auxiliary data.

1.3.1 Errors in Experience

This section presents two key findings. First, potential experience consistently fails to capture actual experience, regardless of the database used to measure the relationship. Second, Black males tend to accumulate experience at a slower rate than do White males, especially for non-college graduates. In establishing these findings, I use three databases. First, the 1993 and 2003 NSCG databases are used to examine college graduates. I establish that traditional potential experience does a poor job of measuring years since graduation for college graduates in the Census. By construction, this is due entirely to differences in the average age at which White and Black males obtain a Bachelor's degree. In addition, the two NSCG databases show that Black males accumulate self-reported professional experience at a lower rate on average than do White males.

Since the NSCG interviews only college graduates, the data contained in that survey are not used as part of the main analysis of this paper, which requires information about all educational attainment levels. However, as these individuals were interviewed in the Census three years prior to being interviewed for the NSCG, the survey can be used to establish that the use of potential experience in the Census has significant drawbacks for the Census sample. In additional to potential experience (defined as $age - 22$ for individuals with a BA/BS), the NSCG has two variables of interest: the age at which individuals obtained their Bachelor's degrees, and self-reported part-time and full-time years of professional experience. The former can be used to construct a refined potential experience measure, and the latter can be used to give an estimate of experience that the individuals consider applicable to their current jobs.

Tables 1.3 and 1.4 show the average years since graduation for Black and White males in the 1993 NSCG and 2003 NSCG. Since potential experience attempts to measure possible out-of-school working years, years since graduation is a more refined version of this variable. The last column of Table 1.3 illustrates the main problem with using a single potential experience measure for each race in 1993: the more accurate measure of potential experience results in a measure that varies systematically by race and across time. For example, White males born between 1934 and 1937 obtained their college degree over four years sooner, on average, than did comparable Black males. Note, however, that for cohorts born between 1962 and 1964, this difference has fallen to less than a year. The last column of Table 1.4 shows similar results for 2003: older cohorts have average differences of over 2 years, while younger cohorts have average differences that are often well less than one half-year. In both databases the difference between the average age at which White and Black males obtain their Bachelor's degrees is larger for older cohorts.

Table 1.5 shows the average years of full-time and part-time professional experience using the 1993 NSCG (professional experience measures are not available in the 2003 NSCG). Black males tend to have more years of part-time professional experience and fewer years of full-time professional experience than their White counterparts. The net effect is that, in general, Black males have fewer total years of professional work experience than Whites.

In addition to the NSCG, I use the NLSY to establish the relationship between accumulated experience and potential experience for individuals present in that survey. Unlike the NSCG, these individuals have not necessarily been interviewed in the Census, and only a narrow band of birth cohorts is available in the NLSY: all individuals in the NLSY are born between 1957 and 1965. The PSID is used in a manner similar to that of the NLSY, and leads to qualitatively similar findings. The use of the PSID in addition to the NLSY adds an additional sample to the analysis, as well as different birth cohorts.

The NLSY contains alternative measures of accumulated experience. Following Regan and Oaxaca (2009), I calculate actual years of experience as the accumulated lifetime total hours worked by each individual for each sample year divided by 2080. This gives a measure of the Full-Time-equivalent years worked. An appealing characteristic of hours worked in the NLSY is that the survey question asks for the respondent's hours worked since their last interview. Thus, if a respondent does not interview for two years, and then re-enters the sample, he should report total hours worked during the previous three years. This can be used to calculate average hours worked for each year that the respondent is not in the survey sample. When calculating the probabilities for the MLE method, years of experience are rounded to the nearest year for tractability reasons (for both the NLSY and the PSID).

Figure 1.1 displays key differences in the cumulative hours worked experience measure in the NLSY by education level. Two implications of this figure stand out. First, the actual experience gap between the races widens as potential experience increases. The accumulation of actual experience for each year of potential experience is lower for Blacks than for Whites, and is persistent. Second, this gap appears to shrink as educational attainment rises. The difference between college graduates and high school dropouts is especially marked. Black high school dropouts with roughly 30 years of potential experience appear, on average, to have accrued roughly seven fewer years of actual experience than do their White cohorts. Conversely, the difference in projected accumulated experience between Black and White college graduates is negligible.

To obtain an experience mapping from the PSID, I compute cumulative years worked (via hours worked) for a subset of individuals. Because of the nature of the hours worked variable ("How many hours have you worked in the past year?"), I focus on individuals that I observe for the vast majority of their working life, with few to no breaks in survey response. To obtain this sample, I look at individuals who first enter the survey at an age younger than twenty. I assume that individuals may begin working at sixteen years of age. Since individuals (especially high school dropouts) may have accumulated work experience by age seventeen, I impute the likely hours worked at age sixteen for individuals who enter the sample between the ages of seventeen and nineteen.[11]

Because of the nature of the question about hours worked in the PSID, once an individual is not interviewed in a survey year, he is dropped from the sample used in this analysis. This is because, unlike the NLSY79 which asks for hours worked since the last survey, the PSID asks only for hours worked in the last year. Thus, once a person drops out of the survey, even temporarily, I am unable to recover the number of hours worked for the missing years. Further, once the survey

[11]For individuals who enter at seventeen, I simply regress accumulated experience at age sixteen on hours worked at ages seventeen through nineteen, and then predict experience accumulated hours worked at age sixteen for these individuals. A similar process is followed for individuals who enter the survey between the ages of seventeen and nineteen.

becomes biennial, similar issues arise, and thus I end my analysis at the last year of annual surveys (this is another issue avoided in the NLSY because of the wording of the survey question).

Figures 1.3-1.6 show the distribution (via box-and-whisker plots) of experience, by level of potential experience, for Black and White males across the four educational attainment levels. Similar to Figures 1.1 and 1.2, college graduates seem to accrue experience at a higher rate than do individuals with different levels of experience. Further, the distribution of experience given potential experience is tighter for college graduates than non-college graduates (this is especially stark when comparing college graduates to high school dropouts). Once the appropriate measures of accumulated experience are obtained in each database, the probability that an individual has a specific amount of experience given his education, age and race (the conditional π_{ij}) can be estimated by the method discussed in Section 2.

1.3.2 Errors in Education

In addition to errors inherent in potential experience in the Census, error also exists in reports of educational attainment. One study that looks at this issue specifically for the Census is Black et al. (2003). They link college graduates in the 1990 Census to individuals in the 1993 NSCG (uniquely), and argue that the educational attainment measure in the NSCG is likely more accurate than that in the Census. Using the reported education levels in the Census and NSCG, they are able to measure college degree over-reporting by race, and find that Black individuals in the NSCG who report having a Bachelor's Degree in the Census are just over twice as likely to have no college degree than are Whites in the same category (13.2% compared to 6.4%)[12]. Due to the nature of the survey's focus, however, analyses using the NSCG are limited to individuals who reported having a college degree in the Census.

An analysis of the errors present in self-reported educational attainment suffers from two problems not present in the analysis of the error in potential experience: a lack of a consistently misreported educational attainment variable, and a lack of a convincing validation or "correct" variable. The former can be attributed to different questions and survey methods used in collecting the educational attainment data in the various surveys. In this analysis, educational attainment is assumed to be measured with error to an equal extent across all of the databases used. In reality, the error is likely worse in the Census than in the auxiliary data since random errors are more likely to

[12]See Table 3 in Black et al. (2003).

be caught and undetected intentional misreporting is more costly in a longitudinal dataset. Finally, note that due to sample size limitations (especially in the sibling measure), I assume that education reporting errors do not vary by age.

Measurement error exists in the educational attainment variable in the NLSY. 9.3% of all individuals in the survey sample, at some point between 1979 and 1994, reported having a lower number of years of education than had been reported in the previous year. Note that the NLSY does provide a "revised" version of this variable that has the logical inconsistencies eliminated (for the most part), but the purpose of this exercise is *not* to find the most consistent version of the education report unless it is also likely to be completely accurate. Instead, it is to try to compare error across responses for (essentially) the same variable in two different databases. As such, the unrevised variable is treated as equivalent to reported educational attainment in the Census.

I focus on two potential measures of true educational attainment. First, I construct a validation measure. Cumulative enrollment is used to expose some over-reporting of educational attainment. The second is a measure that looks at the educational attainment level of an individual, as reported by his siblings. I then refine this sibling approach to greatly reduce the random measurement error that results from random error in sibling responses (since siblings are less likely to know an individual's educational attainment level than is the individual himself). Since the predicted conditional probabilities from the multinomial logit model are equal to the empirical conditional probabilities, the discussion below focuses on the observed conditional probabilities for ease of discussion.

Enrollment data is intuitively appealing because it allows the researcher to link educational attainment responses to longitudinally observed years in school. The use of enrollment data allows for the elimination of over-reporting in the educational attainment variables. The enrollment data cannot be used to identify under-reporting, however, since over-enrollment is prevalent in the data, as shown in Table 1.6. In years where individuals are not in the survey sample, a probit model is used that predicts whether an individual was enrolled in school based on whether the individual was enrolled in previous or subsequent periods, whether the individual did not respond in previous or subsequent periods, and the age and race of the individual. The missing data are imputed instead of the missing years being dropped outright because doing the latter is the same as imputing a value of zero in a cumulative variable. That is, if I see someone enrolled in 1991 and 1993, but that person is not in the sample in 1992, dropping 1992 results in two years of cumulative enrollment between

1991 and 1993 — the same as cumulative enrollment would be were a value of zero imputed for 1992. Since, if I want to keep the individuals in the sample at all after the are not in the survey for a year, the cumulative enrollment measure will have a small degree of measurement error, I choose the impute the data using the probit regression to minimize this error. The results of this measure can be seen in Table 1.7. Similar to Black et al. (2003), I find that Black males are more likely over-report their education than are White males, although to a lesser degree than was reported in their paper — the probabilities in Table 1.7 indicate that Black and White males who report having a college degree actually have no degree with a 4.1% and 1.2% probability, respectively. We also see that, using the cumulative enrollment measure, Black males are slightly more likely to over-report having some college and a high-school diploma than are White males.

Another possible measure of educational attainment involves the use of a question in the 1993 NLSY that asks individuals to report the educational attainment of each of their siblings. If these siblings are in the sample as well, then the self-reported educational attainment variable can be compared to all of the sibling reports of that individual's education level. Unlike the enrollment measure, this measure enables observation of both over-reporting and under-reporting of educational attainment. In an attempt to eliminate the random error introduced by the use of sibling reports of and individual's education level (since siblings are less likely to know the education level of a person than the person himself), I look only at individuals who, in 1993, had at least two siblings interviewed in the 1993 NLSY. Further, all of the siblings must agree on the education level of the individual. When this occurred, I assumed that the siblings were correct with probability one, and compared this "correct" educational attainment measure with that reported by the individual in question. Note that, even with a large number of siblings (five, for example), if four siblings reported an individual as having a college degree, and the fifth reported that the individual simply had some college, then the individual in question was excluded from the subsample. Table 1.8 compares this variable to self-reported educational attainment. An issue that arises in the use of this method is the resulting small sample sizes- as Table 1.8 shows, the cell sizes ranged from 156 observations to 17. As with the cumulative enrollment measure, the percentage of individuals who over-report having obtained a college degree (5.8% for Blacks and 2.13% for Whites) is substantially lower than that in Black et al. (2003) (13.2% for Blacks and 6.4% for Whites).

1.4 Results

To estimate the Black-White wage differential, I focus on young (aged 22-44) Black and White males in the 2000 Census. After eliminating individuals who do not meet the gender and age criteria, the sample consists of 1,662,758 individuals, and is described in Table 1.2. Note that there exist negative predicted values for experience and experience squared. Also, Table 1.2 shows that Blacks have consistently lower predicted values of experience and education than do their White counterparts. This contrasts with the results implied by the potential experience variables, which indicate that Black males have as much (if not slightly more) experience on average than do White males. Appendix A shows first-stage results from the Gilleskie and Mroz (2004) smoothing of experience. Appendix B shows the first-stage results for the Multi-Sample 2SLS along with graphs of the predicted values for experience.

The results of the Multi-Sample 2SLS and MLE methods are shown in Tables 1.9-1.11. For many of the estimates, the standard errors tend to be large; however, I will show that all of the point estimates tell a consistent story—the estimated black-white wage differential using reported values is too large. Table 1.9 focuses solely on accounting for the error present in the potential experience variable. Two key results emerge. First, the estimated conditional wage differential falls considerably when either method is used. The most conservative of the four estimates (Multi-Sample 2SLS using PSID experience data) reduces the estimated wage differential to around 20 percentage points, a reduction of well over a third. Other estimates suggest that the real conditional wage differential is as low as 14.6 percent, a reduction of more than a half. A consistent story is told by the four estimates- the estimated wage differential obtained when using potential experience is much too large. Another result, and one consistent with the literature for gender wage gaps, is that shifting from potential experience to predicted experience (or, here, probabilistic experience) results in a reduction of the estimated returns to schooling and in increase in the returns to experience. This is a result to be expected given the general trend observed in Figures 1 and 2: as educational level increases, the amount of accrued experience per year of potential experience increases, especially for Blacks. Thus, when potential experience is used, some of the returns that are due to higher levels of actual experience are incorrectly attributed to education instead, as a higher level of education will imply a high level of accumulated work experience which will not be captured by the use of potential experience.

Table 1.10 assumes that potential experience accurately measures accumulated experience, and that only educational attainment is reported with error. Correcting for solely the error in education[13] results in a substantially smaller impact on the wage differential than does correcting for solely the error in experience. The *greatest* reduction in the conditional wage differential is 5 percentage points, as opposed to the 14 percentage point *minimum* reduction obtained by correcting solely on experience, found in Table 1.9. Even though the reductions in the estimates of the wage differential are small, the use of predicted or probabilistic education results in point estimates that are consistently smaller (in absolute value) than that provided by OLS.

Table 1.11 assumes errors exist in both potential experience and the reports of educational attainment. When controlling for both mismeasured education and experience, the results suggest that the gains from correcting for error in education and potential experience are approximately additive. Using sibling reports of education to generate predicted and probabilistic education measures results in a greater reduction in the wage differential than does using the cumulative enrollment variable. This is not unexpected, since the cumulative enrollment measure is, at least in one dimension, inferior to the sibling reports measure given that the latter can identify both over-reporting and under-reporting while the former is limited to identifying over-reporting. Regardless of the measure used, the largest (in absolute value) estimated conditional wage differential is 20 percent. The smallest estimated differential is 9 percent. These results indicate that the use of potential experience and reported education in a Census Mincer-style regression causes the estimated racial wage differential to be overstated by between 14.5 and 25.5 percentage points.

The error also impacts other estimates. The results in Tables 1.9 and 1.11 show that correcting for the error in education unilaterally decreases the estimated return to education obtained from the uncorrected OLS. The reason for this decrease can be inferred from Figures 1.1 and 1.1: the estimated returns to increasing education will no long also include the returns to higher levels of accumulated experience that seem to accompany this education increase, but which cannot be detected when using potential experience. Further, we see differences in the estimated returns (marginal effects) to eduction when we use the corrected measures. As an example, Figure 1.7 shows the estimated returns to experience when we use the NLSY to correct experience and the sibling measure to correct education. The MLE estimates essentailly show an upward vertical shift in the returns to experience of roughly 3%. The estimated returns to experience from the Multi-

[13]The ML method adjusts potential experience appropriately for all possible educational attainment categories

Sample 2SLS estimates, on the other hand, represent a significant departure from the shape of the returns-to-experience curve suggested by OLS — the curve estimated by Multi-Sample 2SLS has higher returns to additional experience when the stock of experience is low, but the decline is much steeper.

These results also highlight a significant advantage of using the Maximum Likelihood method instead of Multi-Sample 2SLS: when the errors in experience are accounted for, the standard errors of the ML estimates are substantially smaller than are those of Multi-Sample 2SLS, especially when the PSID auxiliary data is used. In fact, when PSID data is used, we cannot make meaningful inference about the size of the Black-White wage differential, as the 95% confidence interval covers a range of about 80 percentage points! This difference is evident for other estimates as well, and is particularly prominent in the standard errors of the experience variables. The standard errors exceed the coefficients for EXP and EXP^2 for all of the Multi-Sample 2SLS estimates that correct for experience, but are generally no greater than about half of the value of the coefficients in the MLE specifications.

1.5 Conclusion

Measurement error inherent in potential experience and reported education has a real and substantial impact on estimates of the Black-White wage differential. I show this in the 2000 Census first by estimating the racial wage differential using potential experience and reported education, and comparing these results to those found when different measures of predicted or probabilistic education and experience measures are used. The results are stark: the racial wage gap falls from an estimated 34.5 percent to somewhere between 9 and 20 percent. This reduction is due purely to correcting for measurement error, as the remaining independent variables are unchanged.

Use of potential experience and reported education causes an inflated wage differential due to the nature of the measurement error in each. Potential experience, while appearing to be a reasonable proxy for White males, overstates the accumulated experience for Black males. The magnitude of the error increases across time, and it is especially pronounced for non-college-graduates. For example, a Black high school dropout with 30 years of potential experience has, on average, seven fewer years of accumulated experience than does a White high school dropout, as illustrated in Figure 1.1. Further, Black males seem more likely to overstate their level of educational attainment than do comparable

White males. Each of these results in part of the differential that is actually due to education and experience to be attributed to part of the "unexplained" Black-White wage differential. The fact that the accumulated experience gap between Black and White males narrows as educational attainment increases makes estimation of a Mincer-style wage equation more problematic: not only is the racial wage gap estimate biased, but the educational attainment and experience effects are biased as well.

Since the distribution of measurement error can be reasonably approximated due to the existence of auxiliary data, I introduce an estimation technique that assists with estimation of the results discussed above. This Maximum Likelihood estimation method performs well in a simulation study, and directly integrates the distribution of the measurement error into the estimation procedure.

In addition to the introduction of the Maximum Likelihood estimation technique, the results of this paper add to the existing literature in several ways. First, it integrates much of the knowledge gained concerning measurement error in potential experience in estimates of the gender wage gap to estimates of the racial wage gap in a more rigorous manner than has been done previously. Second, it informs researchers using Census data of better ways to control for experience and education in wage regressions. Finally, it also shows that using potential experience in situations where predicted or accumulated experience are available should be avoided, as using these measures is likely to lead to severely biased estimates.

1.6 Tables and Figures

Figure 1.1: Relationship between Cumulative Work Hours and "Potential" Experience: Evidence from the NLSY

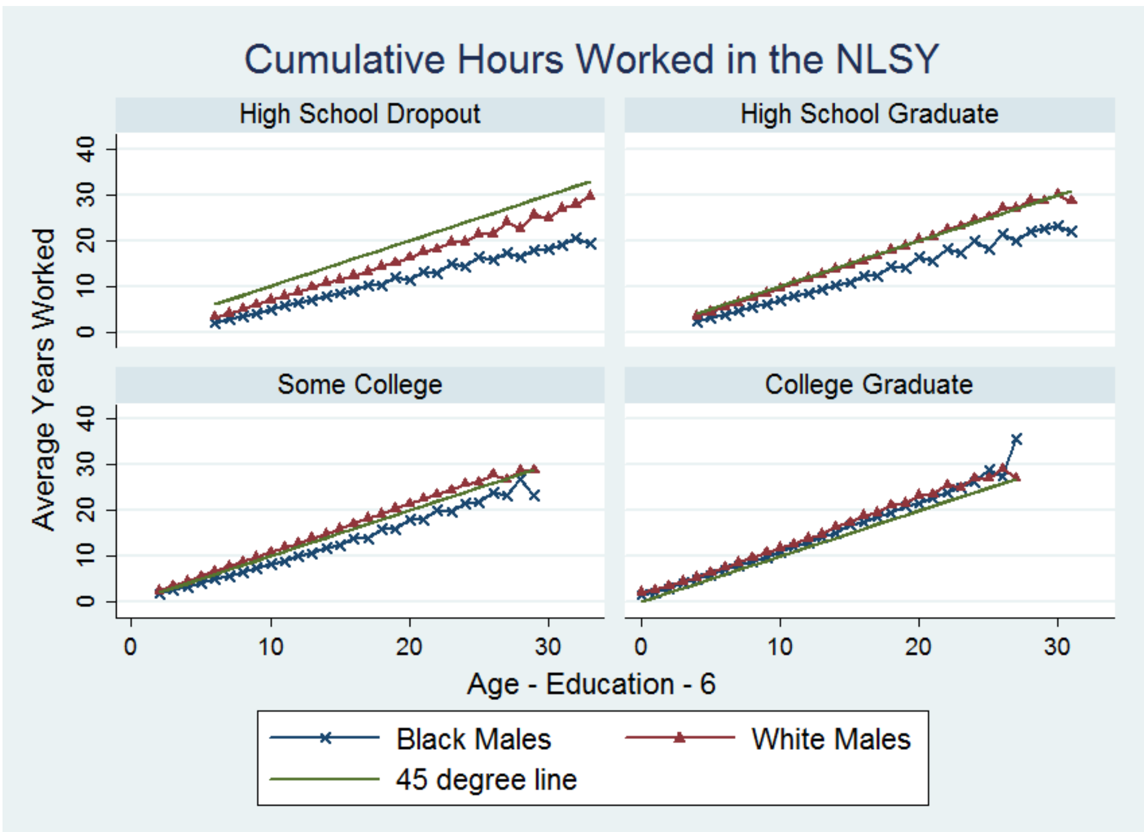

Figure 1.2: Relationship between Cumulative Work Hours and "Potential" Experience: Evidence from the PSID

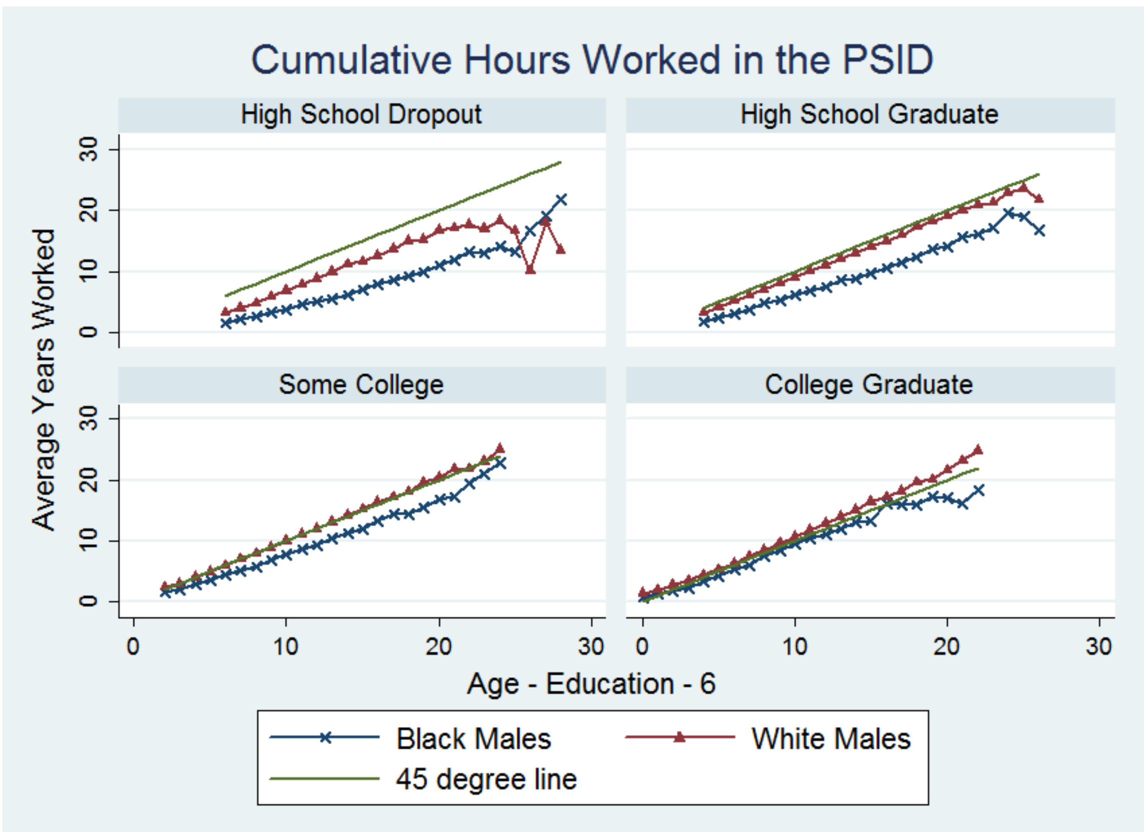

Figure 1.3: Relationship between Cumulative Work Hours and “Potential” Experience of High School Dropouts

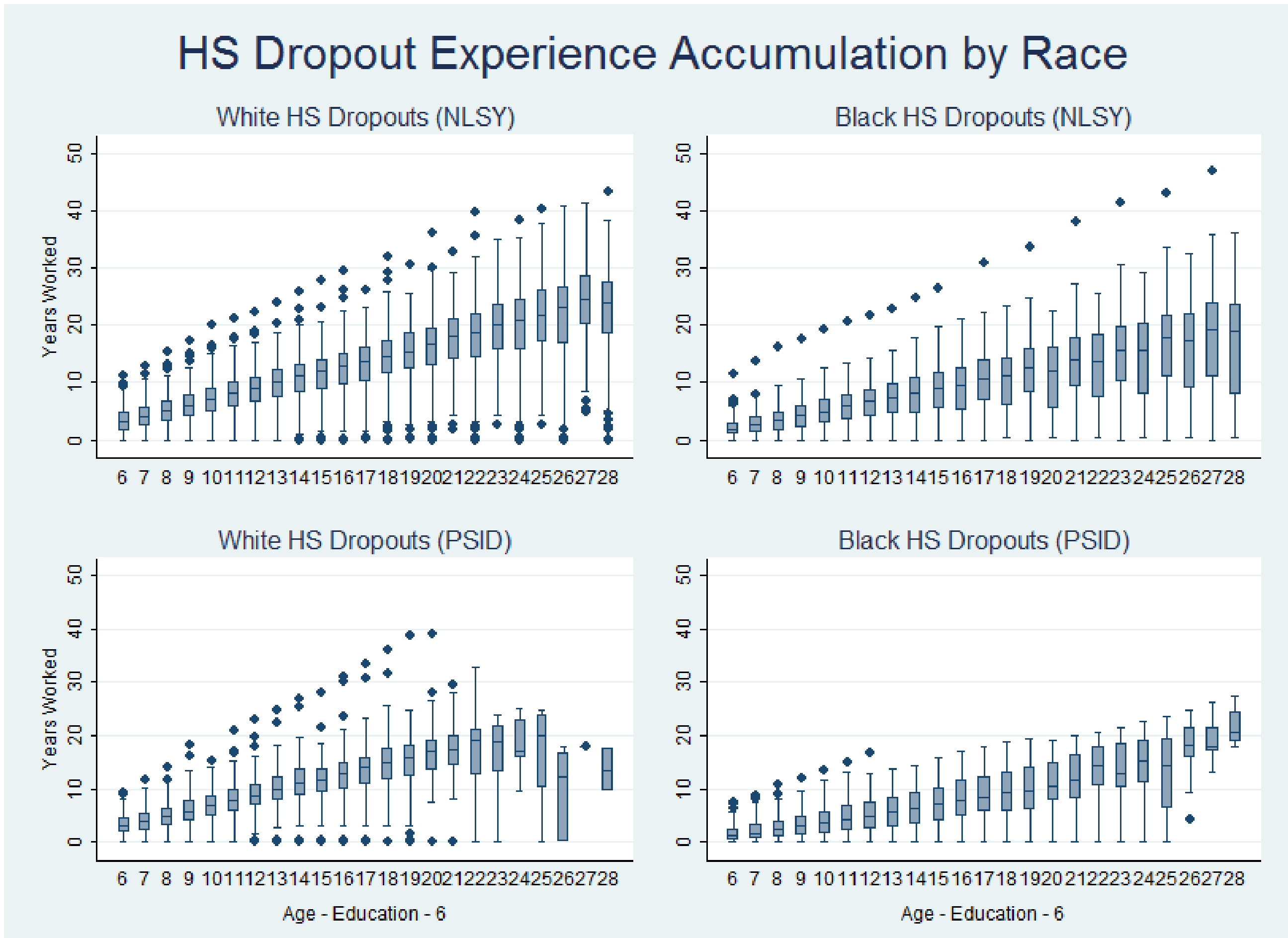

Figure 1.4: Relationship between Cumulative Work Hours and "Potential" Experience of High School Graduates

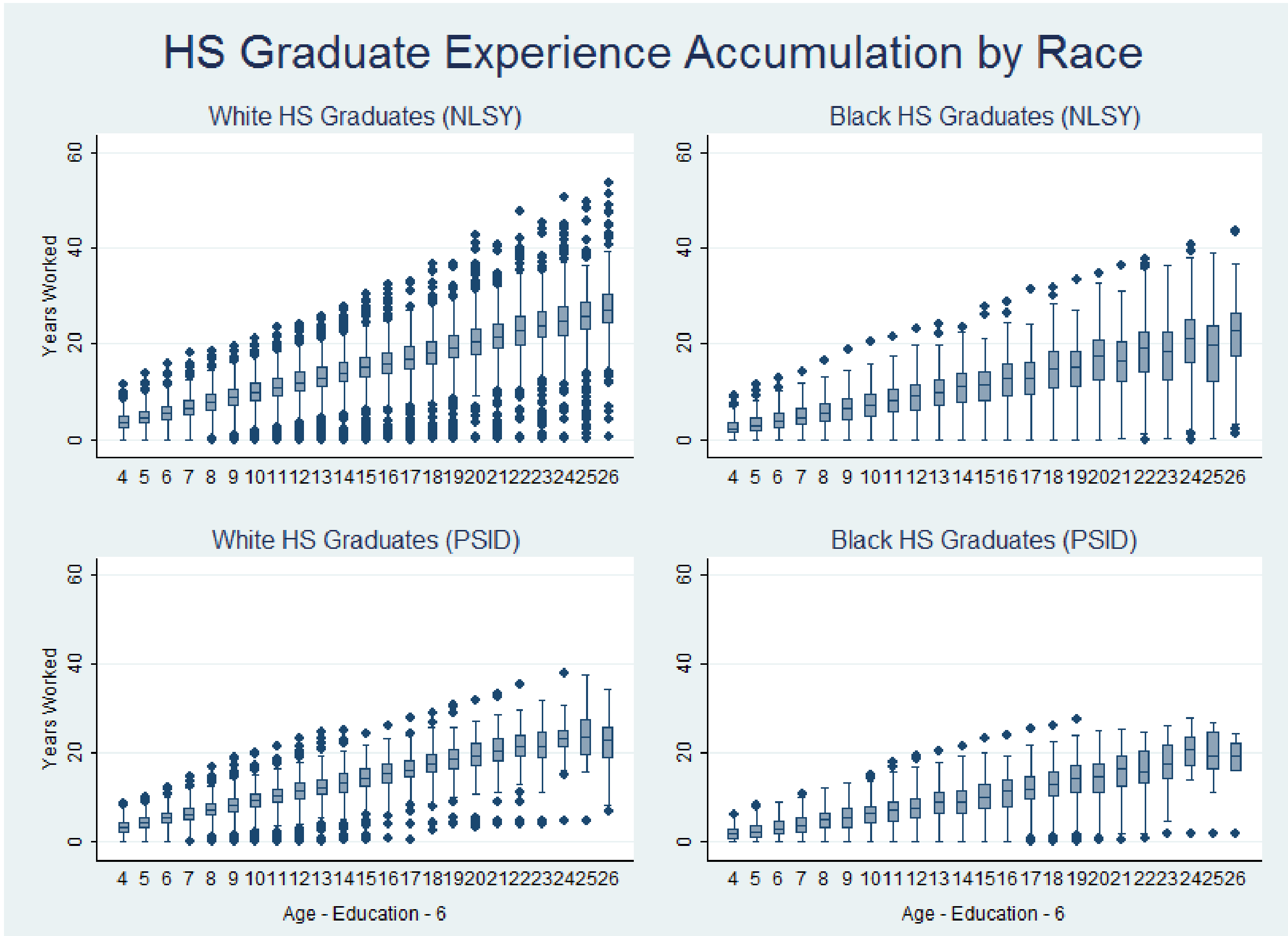

Figure 1.5: Relationship between Cumulative Work Hours and "Potential" Experience of Individuals with Some College

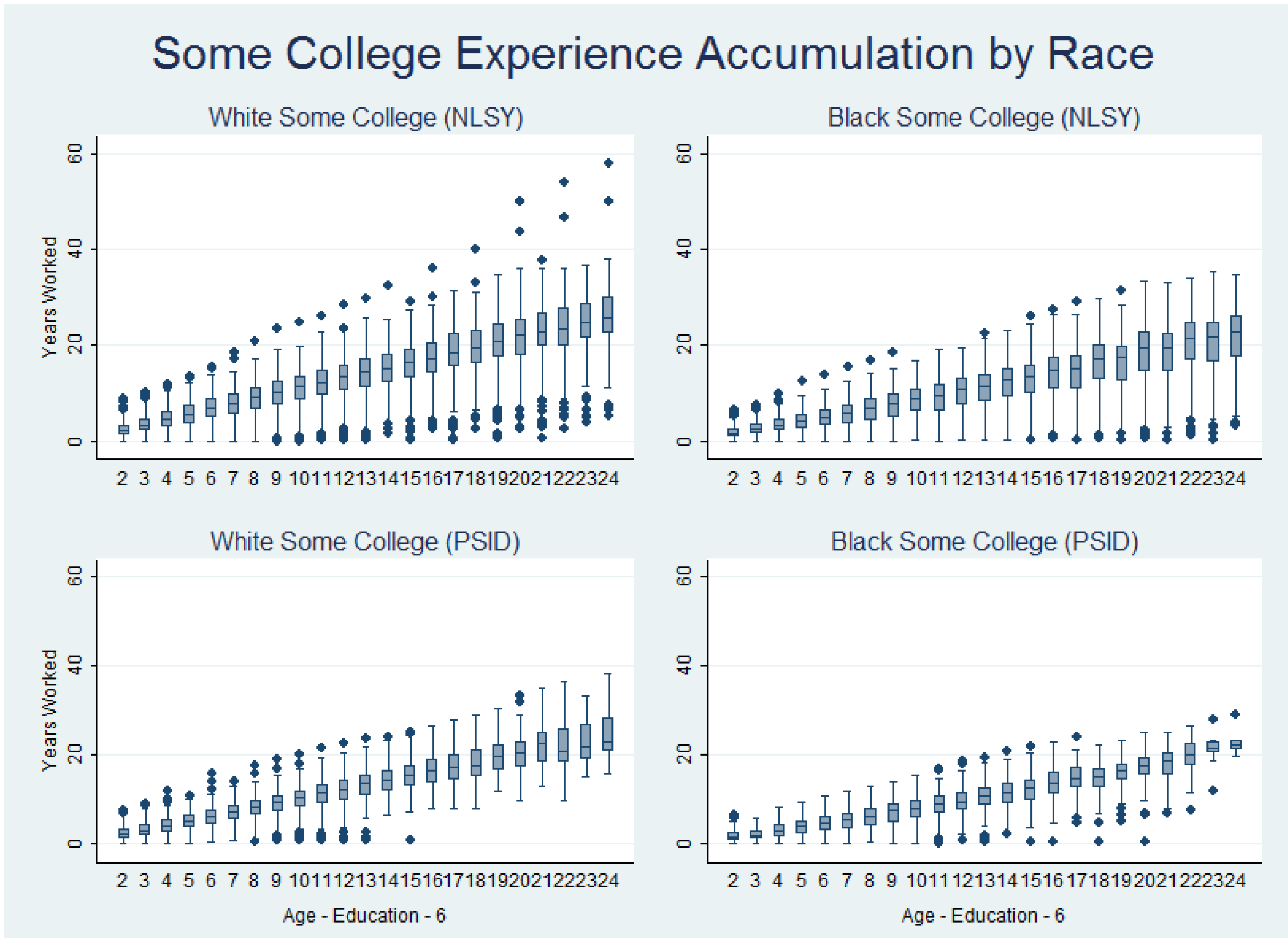

Figure 1.6: Relationship between Cumulative Work Hours and “Potential” Experience of College Graduates

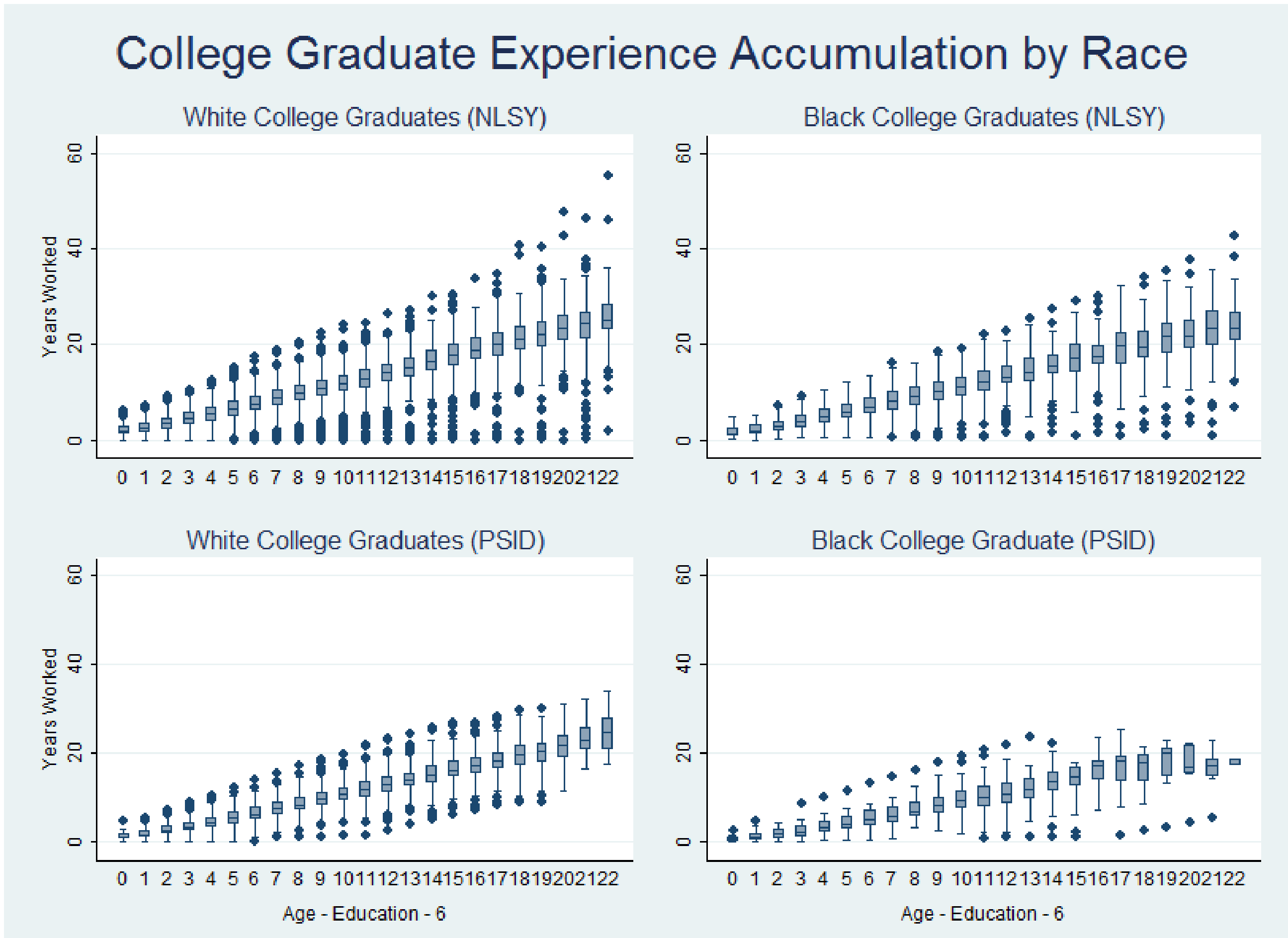

Figure 1.7: Returns to Experience: Comparing OLS, MLE, and Multi-Sample 2SLS

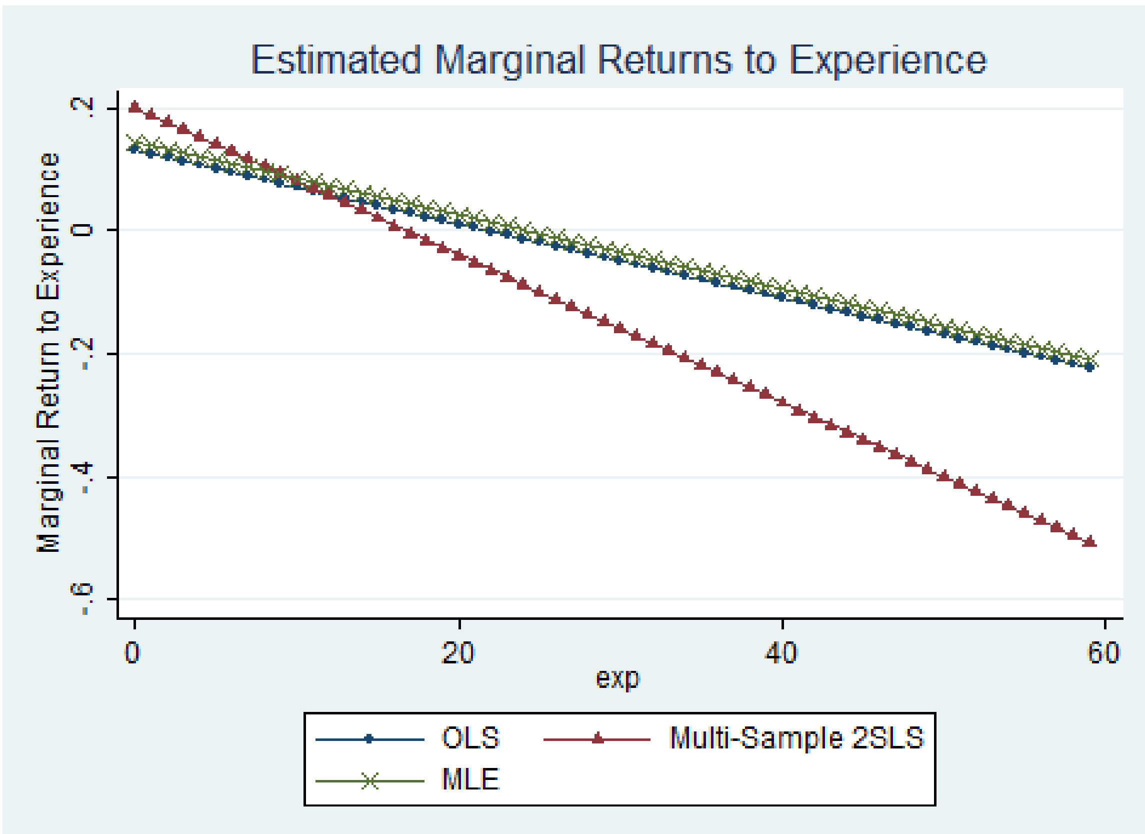

Table 1.1: Evaluation of Estimation Methods using Simulation

	Truth	OLS with True Values	OLS with Reported Values	Regan & Oaxaca Two-Stage	Multi-Sample 2SLS	Maximum Likelihood
Black	−0.2	−0.200	−0.280	−0.213	−0.201	−0.196
σ		(0.032)	(0.037)	(0.042)	(0.042)	(0.037)
Education	0.15	0.150	0.168	0.143	0.150	0.147
σ		(0.005)	(0.006)	(0.007)	(0.007)	(0.006)
Experience	0.15	0.150	0.118	0.222	0.151	0.146
σ		(0.004)	(0.007)	(0.012)	(0.012)	(0.007)
Experience^2	−0.0025	−0.0025	−0.0020	−0.0062	−0.0025	−0.0024
σ		(0.0001)	(0.0003)	(0.0005)	(0.0004)	(0.0002)
Constant	7	7.00	6.68	6.73	7.00	7.04
σ		(0.072)	(0.103)	(0.116)	(0.119)	(0.097)

This table displays the results of 1,000 simulations in databases with 10,000 observations.

Table 1.2: Census Summary Statistics

Variable	Mean (σ)	Mean (σ)
	White (N=1,472,555)	**Black (N=190,203)**
Log Wages	10.3 (0.9)	9.9 (1.0)
Age	34.0 (6.5)	33.4 (6.5)
Reported Education	13.4 (1.9)	12.8 (1.7)
Predicted Education (Sibling)	13.6 (1.7)	12.8 (1.6)
Predicted Education (Enroll)	13.4 (1.9)	12.7 (1.7)
Potential Experience	14.6 (6.7)	14.6 (6.7)
Predicted Experience (NLSY)	14.8 (6.2)	11.7 (6.2)
Predicted Experience (PSID)	13.7 (5.8)	10.6 (5.8)

Table 1.3: Evaluation of Potential Experience in the 1993 NSCG

			Years Since Graduation		
Birthyear	Age	Potential Experience	White	Black	Black − White
1929	64	42	38.10	35.32	−2.79
1930	63	41	37.32	32.77	−4.55
1931	62	40	36.36	32.92	−3.44
1932	61	39	35.31	31.59	−3.72
1933	60	38	34.25	31.86	−2.39
1934	59	37	33.16	28.30	−4.86
1935	58	36	32.62	28.22	−4.40
1936	57	35	31.60	27.56	−4.04
1937	56	34	30.42	26.38	−4.05
1938	55	33	29.64	25.68	−3.96
1939	54	32	28.40	26.52	−1.88
1940	53	31	27.81	22.69	−5.12
1941	52	30	26.92	24.72	−2.20
1942	51	29	26.13	23.43	−2.70
1943	50	28	25.39	23.24	−2.15
1944	49	27	24.48	21.38	−3.10
1945	48	26	23.52	21.25	−2.27
1946	47	25	22.42	21.59	−0.83
1947	46	24	21.68	19.67	−2.00
1948	45	23	20.74	18.60	−2.14
1949	44	22	19.89	18.65	−1.25
1950	43	21	19.12	17.67	−1.45
1951	42	20	18.16	17.22	−0.94
1952	41	19	17.04	16.70	−0.34
1953	40	18	16.17	15.23	−0.94
1954	39	17	15.29	14.15	−1.14
1955	38	16	14.34	13.42	−0.92
1956	37	15	13.30	12.62	−0.67
1957	36	14	12.48	12.14	−0.35
1958	35	13	11.73	10.78	−0.95
1959	34	12	10.70	10.24	−0.47
1960	33	11	9.71	9.39	−0.32
1961	32	10	8.85	8.38	−0.48
1962	31	9	7.95	7.68	−0.27
1963	30	8	7.00	6.84	−0.16
1964	29	7	6.16	6.02	−0.15
1965	28	6	5.26	5.34	0.07
1966	27	5	4.56	4.61	0.05

Table 1.4: Evaluation of Potential Experience in the 2003 NSCG

			Years Since Graduation		
Birthyear	Age	Potential Experience	White	Black	Black − White
1939	64	42	38.51	33.99	−4.52
1940	63	41	37.84	34.24	−3.61
1941	62	40	36.71	33.16	−3.55
1942	61	39	36.21	31.21	−5.00
1943	60	38	34.76	32.68	−2.08
1944	59	37	33.85	30.45	−3.40
1945	58	36	33.00	27.73	−5.27
1946	57	35	31.77	29.09	−2.69
1947	56	34	31.43	30.62	−0.81
1948	55	33	30.23	28.69	−1.55
1949	54	32	29.41	28.40	−1.01
1950	53	31	28.68	27.43	−1.25
1951	52	30	27.15	23.98	−3.17
1952	51	29	26.17	24.32	−1.84
1953	50	28	25.31	22.37	−2.95
1954	49	27	24.43	22.25	−2.19
1955	48	26	23.17	21.89	−1.28
1956	47	25	22.31	20.53	−1.78
1957	46	24	21.49	19.56	−1.93
1958	45	23	20.45	19.00	−1.45
1959	44	22	19.59	18.38	−1.21
1960	43	21	18.80	16.99	−1.81
1961	42	20	17.51	15.79	−1.73
1962	41	19	16.90	15.45	−1.45
1963	40	18	15.83	14.68	−1.15
1964	39	17	14.88	14.16	−0.72
1965	38	16	13.89	13.37	−0.52
1966	37	15	13.29	12.70	−0.58
1967	36	14	12.09	11.31	−0.78
1968	35	13	11.27	11.02	−0.24
1969	34	12	10.34	9.37	−0.97
1970	33	11	9.46	9.13	−0.33
1971	32	10	8.58	8.28	−0.29
1972	31	9	7.73	6.69	−1.04
1973	30	8	6.85	6.54	−0.30
1974	29	7	5.94	5.80	−0.14
1975	28	6	5.31	5.25	−0.06
1976	27	5	4.58	4.57	−0.01

Table 1.5: Years of Professional Experience in the 1993 NSCG

Birth Year	Age	Potential Experience	White Experience	Black Experience	Black − White	
					Full Time Only	Total Experience
1929	64	42	33.71	29.51	−3.88	−4.20
1930	63	41	32.43	25.37	−6.89	−7.05
1931	62	40	32.89	29.61	−2.88	−3.28
1932	61	39	32.91	28.87	−4.55	−4.04
1933	60	38	31.37	27.19	−4.87	−4.17
1934	59	37	31.15	25.64	−6.32	−5.51
1935	58	36	31.44	28.75	−3.63	−2.69
1936	57	35	29.38	25.68	−4.45	−3.71
1937	56	34	28.69	27.25	−1.63	−1.44
1938	55	33	28.35	25.40	−3.42	−2.95
1939	54	32	27.20	25.92	−2.52	−1.29
1940	53	31	27.36	24.89	−3.58	−2.47
1941	52	30	25.73	23.72	−2.29	−2.01
1942	51	29	25.32	24.29	−0.99	−1.04
1943	50	28	24.55	20.61	−3.66	−3.94
1944	49	27	23.32	21.34	−2.93	−1.98
1945	48	26	21.90	21.41	−0.88	−0.48
1946	47	25	21.18	19.77	−1.45	−1.41
1947	46	24	20.53	19.22	−1.41	−1.31
1948	45	23	19.66	17.41	−2.69	−2.25
1949	44	22	18.60	18.61	−0.30	0.01
1950	43	21	18.00	17.85	−0.35	−0.15
1951	42	20	17.21	15.47	−1.71	−1.74
1952	41	19	16.15	15.77	−0.34	−0.38
1953	40	18	15.42	14.18	−1.32	−1.25
1954	39	17	15.13	13.71	−1.57	−1.42
1955	38	16	14.05	12.80	−1.19	−1.25
1956	37	15	13.33	12.73	−1.20	−0.60
1957	36	14	12.31	11.62	−0.71	−0.69
1958	35	13	11.45	11.21	−0.39	−0.24
1959	34	12	10.67	10.47	−0.47	−0.20
1960	33	11	9.77	9.54	−0.58	−0.23
1961	32	10	9.01	8.80	−0.28	−0.21
1962	31	9	8.25	7.98	−0.26	−0.27
1963	30	8	7.40	6.98	−0.28	−0.42
1964	29	7	6.72	6.74	−0.19	0.02
1965	28	6	5.81	5.85	−0.13	0.04
1966	27	5	5.11	4.82	0.11	−0.29

Here, experience is defined as years of full-time plus one-half years of part-time professional experience.

Table 1.6: Enrollment and Reported Education in the NLSY, Percent Agreement

	White Reported Education Group				Black Reported Education Group			
	HS Dropout	HS Grad	Some College	College Grad	HS Dropout	HS Grad	Some College	College Grad
Years enrolled								
<9	20.81	0.54	0	0	13.99	0.75	0	0
9	20.97	1.25	0	0	16.51	0.97	0	0
10	22.93	2.74	0.2	0	25.52	3.72	0.18	0
11	23.55	5.42	0.63	0	26.76	6.83	0	0
12	7.65	69.91	0.49	0	9.86	62.37	2.35	0
13	1.94	11.89	13.67	0	4.52	17.13	14.77	1.52
14	0.2	5.25	24.37	0	0.89	5.68	21.55	1.52
15	0.53	1.82	24.51	1.2	0.46	1.46	21.8	1.06
16	0.28	0.7	15.77	31.98	0.5	0.74	17.91	23.08
17	0.31	0.27	9.13	24.78	0.46	0.09	9.67	26.8
18	0.26	0.07	4.92	14.64	0.07	0.19	5.24	17.77
19	0.13	0.06	3.18	10.84	0.21	0.07	2.89	9.72
20	0.03	0.03	1.54	8.11	0.04	0	1.53	6.76
21	0.08	0.01	0.56	4.43	0.14	0	0.78	5.47
>21	0.33	0.03	1.05	4.02	0.07	0	1.32	6.3
N:	3911	13246	5917	6497	2810	6782	2803	1317

Table 1.7: Education and Enrollment-Corrected Education the NLSY, Percent Agreement

	White Reported Education Group				Black Reported Education Group			
	HS Dropout	HS Grad	Some College	College Grad	HS Dropout	HS Grad	Some College	College Grad
Education from enrollment								
HS Dropout	100	9.96	0.83	0	100	12.27	0.18	0
HS Grad	0	90.04	0.49	0	0	87.73	2.35	0
Some College	0	0	98.68	1.2	0	0	97.47	4.1
College Grad	0	0	0	98.8	0	0	0	95.9

Table 1.8: Education and Sibling-Corrected Education the NLSY, Percent Agreement

	White Reported Education Group				Black Reported Education Group			
	HS Dropout	HS Grad	Some College	College Grad	HS Dropout	HS Grad	Some College	College Grad
Education from siblings								
HS Dropout	100	4.2	0	0	86.67	14.1	0	0
HS Grad	0	89.51	28.57	0	11.11	83.33	57.69	0
Some College	0	4.2	69.39	2.13	2.22	2.56	42.31	5.88
College Grad	0	2.1	2.04	97.87	0	0	0	94.12
N:	21	143	49	94	45	156	26	17

Table 1.9: Results: Correcting only for Experience Error

	No Correction	Multi-Sample 2SLS		MLE	
		PSID exp	NLSY exp	PSID exp	NLSY exp
Black	−0.345	−0.146	−0.202	−0.185	−0.186
	(0.002)	(0.206)	(0.057)	(0.092)	(0.013)
Education	0.169	0.117	0.120	0.137	0.130
	(0.0004)	(0.0790)	(0.0381)	(0.0071)	(0.0020)
Experience	0.130	0.159	0.161	0.128	0.145
	(0.0004)	(0.3281)	(0.2454)	(0.0132)	(0.0055)
Experience^2	−0.003	−0.005	−0.004	−0.003	−0.003
	(0.00001)	(0.01323)	(0.00883)	(0.00024)	(0.00012)
Constant	6.965	7.621	7.515	7.314	7.392
	(0.006)	(0.680)	(0.645)	(0.182)	(0.047)
LLF	−2018844	−2018844	−2018844	−1963833	−1950777

Standard errors are reported in parentheses. “No correction” estimation controls for potential experience and reported education.

Table 1.10: Results: Correcting only for Education Error

	No Correction	**Multi-Sample 2SLS**		**MLE**	
		Enrollment	Sibling	Enrollment	Sibling
Black	−0.345	−0.340	−0.295	−0.339	−0.302
	(0.002)	(0.002)	(0.012)	(0.008)	(0.058)
Education	0.169	0.167	0.188	0.168	0.176
	(0.0004)	(0.0004)	(0.0044)	(0.0027)	(0.0046)
Experience	0.130	0.132	0.171	0.132	0.132
	(0.0004)	(0.0005)	(0.0137)	(0.0062)	(0.0348)
Experience^2	−0.003	−0.003	−0.005	−0.003	−0.003
	(0.00001)	(0.00002)	(0.00052)	(0.00019)	(0.00104)
Constant	6.965	7.002	6.453	6.973	6.871
	(0.006)	(0.006)	(0.084)	(0.009)	(0.317)
LLF	−2018844	−2018844	−2018844	−2018131	−2017232

Standard errors are reported in parentheses. "No correction" estimation controls for potential experience and reported education.

Table 1.11: Results: Correcting for Both Education and Experience Errors

		Correcting with Cumulative Enrollment				Correcting with Sibling Reports of Education			
	No Correction	**Multi-Sample 2SLS**		**MLE**		**Multi-Sample 2SLS**		**MLE**	
		PSID Exp	NLSY Exp	PSID Exp	NLSY Exp	PSID Exp	NLSY Exp	PSID Exp	NLSY Exp
Black	−0.345	−0.143	−0.200	−0.174	−0.179	−0.090	−0.157	−0.134	−0.141
	(0.002)	(0.205)	(0.058)	(0.118)	(0.018)	(0.201)	(0.069)	(0.120)	(0.078)
Education	0.169	0.116	0.119	0.135	0.127	0.123	0.129	0.144	0.136
	(0.0004)	(0.0743)	(0.0352)	(0.0037)	(0.0024)	(0.0357)	(0.0089)	(0.0051)	(0.0253)
Experience	0.130	0.161	0.163	0.128	0.146	0.195	0.200	0.128	0.145
	(0.0004)	(0.3251)	(0.2438)	(0.0140)	(0.0075)	(0.2954)	(0.2312)	(0.0183)	(0.0631)
Experience^2	−0.003	−0.005	−0.004	−0.003	−0.003	−0.006	−0.006	−0.003	−0.003
	(0.00001)	(0.01309)	(0.00876)	(0.00027)	(0.00017)	(0.01191)	(0.00833)	(0.00034)	(0.00128)
Constant	6.965	7.644	7.532	7.232	7.435	7.360	7.179	7.370	6.316
	(0.006)	(0.428)	(0.680)	(0.189)	(0.066)	(0.809)	(0.979)	(0.250)	(0.337)
LLF	−2018844	−2018844	−2018844	−1959455	−1949644	−2018844	−2018844	−1955972	−1947266

Standard errors are reported in parentheses. “No correction” estimation controls for potential experience and reported education.

Chapter 2

The Effect of Market Structure on the Airfare Response to the Entry of Southwest Airlines

2.1 Introduction

The "Southwest Effect" is a term used in the airline literature to describe the large airfare reductions that occur in response to competition from Southwest Airlines. Studies estimating this effect have ranged from case-study approaches (e.g., Dresner et al., 1996; Vowles, 2001; Bennett and Craun, 1993) to more general analyses incorporating a broader range of entry occurrences (e.g., Morrison, 2001; Goolsbee and Syverson, 2008). The consensus in this literature is that the effect of competition from Southwest is substantial, both in the routes in which entry actually occurs *and* nearby in routes that are likely to compete for passengers with the entered route. What is missing from the literature is an analysis of the extent to which the fare effects of Southwest entry varies across routes, as well as the identification of factors that explain this variation in effects. This paper addresses this gap in the literature.

My study is built around the repeal of the Wright Amendment, a natural experiment that allowed for flights out of a major Southwest hub, Love Field, to a range of previously prohibited destinations. I use data on flights to and from Dallas-Fort Worth International Airport to conduct two separate analyses. First, I treat each entered route as a separate case study and use a difference-in-difference analysis to uncover the causal effects of entry by Southwest. I find a median route-level reduction in fares of 11.2%. The route-level estimates vary greatly, however, from a fall of over 40% to a roughly 2.5% *increase* in average fares. Second, I perform a fixed-effects analysis that uses market characteristics to explain the variation in fare effects uncovered by the first analysis. I find that the average airport presence of the ticketing airline and the presence of an existing low-cost carrier are the most important determinants of the size of the Southwest Effect.

For all of the routes in my analysis, my discussion of Southwest entry is focused on entry into a competing (nearby) route. These routes are important units of analysis for several reasons. First, when operating out of multi-airport cities, Southwest often operates out of the smaller (less congested) airport. Second, while Morrison (2001) shows an average fare reduction of 46.2% on directly entered routes, he still finds an economically and statistically significant effect on two types of nearby routes (an average fare reduction of 26.4% routes with one common endpoint and one nearby endpoint, and an average fare reduction of 12.4% on routes with two nearby endpoints). Finally, Goolsbee and Syverson (2008) show that while the threat of entry by Southwest reduces fares before entry actually occurs in directly entered routes, they find no evidence of this effect on nearby routes. Their findings support the use of the Wright Amendment as a natural experiement, as even if the time between the partial and full repeal of the Wright Amendment was enough for airlines operating out of threatened routes to make strategic pricing decisions, these strategic actions were (according to their findings) unlikely to have actually occurred.

2.1.1 The Wright Amendment

The Wright Amendment was legislation introduced shortly before the deregulation of the airline industry in the late 1970s that barred flights between Love Field in Dallas, Texas and airports non-neighboring states.[1] Southwest Airlines is based out of Love Field, and while Southwest was not a significant force in the initial years of deregulation, by the mid-2000s, the Wright Amendment was a powerful shield from competition from Southwest Airlines for carriers flying out of nearby Dallas-Fort Worth International Airport (DFW). The likely importance of this shield to the profitability

[1]Passengers could legally only fly between Love Field and either airports in Texas or airports in neighboring states. While passengers could theoretically create their own round-trip ticket from Love Field to an airport in a non-neighboring state by booking a flight from Love Field to New Orleans and then from New Orleans to the final destination, airlines flying out of Love Field were prohibited by law from advertising this fact and had to require passengers to change planes. Empirically, few passengers are seen constructing such tickets before the repeal.

on routes to and from DFW (via airfares) is implied in the above literature and will be explicitly analyzed later in this paper.

In the mid-2000s, the Wright Amendment was repealed in two parts. First, an amendment that allowed flights between Love Field and two airports in Missouri was introduced by a Missouri legislator, and ultimately became law in late 2005. Southwest Airlines, American Airlines, Love Field, and Dallas-Fort Worth International airport later mutually introduced a formal repeal of the Wright Amendment. This repeal became law in late 2006, after which Southwest began service between Love Field and an additional twenty previously prohibited destinations. This immediate entry into a large number of routes is one of the factors that separates this study from the rest of the literature. Likely selection into the most profitable routes will result in an average estimated effect of competition from Southwest Airlines that will be higher than what the average effect would be if Southwest could, at the same cost, enter every route in the country or if it entered routes randomly. We expect this to be the case because airplanes are inherently mobile capital — each airplane used on a Love Field route could have been used on not just a different Love Field route, but potentially on any route in the country flown by Southwest. Thus, if we see Southwest Airlines enter a route, that route must have been more profitable than the available alternatives. Further, if we think that the most profitable routes are the ones on which fare markups are the highest, then we would see Southwest enter routes where fares have the most "room to move." However, even given the sample selection problem, the removal of across-region and across-time differences in the entered routes allows for the examination of market-level and airline-level characteristics on the airfare response to competition from Southwest that would otherwise not be supported by the available data. Because of the nature of the repeal, all of the analysis concerns the effect of entry by Southwest on a "competing" route, where competing routes are defined as sharing one endpoint while the other is geographically close. Here, all of the routes share a non-Texas endpoint, while the other endpoints (Love Field for the entered routes and Dallas-Fort Worth International for the analyzed routes) are separated by 11 miles.

2.2 Data

The data used in this paper come from the Department of Transportation's Airline Origin and Destination Survey (DB1B) which is a quarterly 10% random sample of all domestic flights in the United States. My analysis focuses on flights from the first quarter of 1999 to the fourth quarter of 2009. Data is available at the (unique) ticket level, with a variable that indicates the number of passengers that bought a specific ticket at a specific price. So, for example, if 253 passengers bought identical tickets from Tampa to New York with a single stop in Charlotte on the same airline at the same price, then that ticket would appear once in the data with a variable indicating that 253 people purchased the ticket. Information about the tickets includes the origin airport, destination airport, and any intermediate stops. Also included is the fare (price), ticket fare class (e.g., first class), whether the ticket was a roundtrip ticket, the carrier under whose name the ticket was sold, the carrier that actually operated the flight, and the year and quarter in which the flight was operated.

While the data are robust, there do exist a couple of limitations. First, each-way fare on a roundtrip ticket is not recorded, so (following the literature) I divide the total fare by two on roundtrip tickets to proxy for the fare for each one-way portion. I am also unable to observe the specific day, time, or flight number of the flight. While this limits the ability to fully specify the fare equation, it is also a problem faced by the majority of the literature.

I use the data from DB1B in two different forms. First, the case study approach uses ticket-level observations for its difference-in-difference analyses for select routes.[2] The data used in these case studies include the log fare of the ticket and information on the origin and destination of the flight. Each specific origin-destination pair constitutes a different route; so, for example, DFW-MDW and MDW-DFW are considered different routes. Also included are two sets of variables involving the airline for which the ticket was issued. The first set is simply a set of indicator variables that identifies the airline. The second set is a series of airline-specific fare class variables. I also use information on the year and quarter of the flight to identify time periods before and after the routes were affected by the repeal of the Wright Amendment.

As the fixed-effects analysis requires within-group comparisons, I use the individual ticket data to create route-level averages for each airline on the route, such as the log of average fare. I also create two variables that indicate the time periods after the Wright Amendment has been repealed. One variable indicates a post-repeal period on all routes, and another indicates a post-repeal period only on the routes affected by the entry of Southwest Airlines on a competing route. For the general post-repeal variable, the post-repeal period is defined as periods no earlier than the fourth quarter of 2006 for most routes. The exception is routes between DFW and the airports in Missouri, for which

[2]The process by which these routes are chosen is discussed in the next section.

the post-repeal period is defined as any quarter-year after 2005.[3] A "threatened entry" variable is also created for the routes affected by the second stage of the repeal, and is an indicator variable for the time period between the fourth quarters of 2005 and 2006 (when the Wright Amendment had been repealed for only the Missouri routes).

I then create several other route-airline-level and route-level variables. First, I calculate route share by measuring the fraction of total enplanements on a route carried by the airline. Enplanements measure whether someone was on a plane that serviced that route and includes individuals who fly the route as part, but not all, of a ticket. So, an ticket from Tampa to New York with stops in Charlotte and Chicago would be included in the route share calculation of a Charlotte-to-Chicago route. [4] To capture the overall presence of the carrier at the endpoints of a route, I calculate the average airport share of a carrier. To obtain this average, I first calculate the fraction of total enplanements (both arriving and departing) at an airport attributable to a specific carrier. The carrier's airport shares at the origin and destination routes are then averaged to create the average share variable. Additional variables include the percentage of passengers flown by low-cost carriers (other than Southwest Airlines) the route, and whether the carrier is a low-cost carrier.

Finally, I create three route-level variables that capture route characteristics widely used in the literature. The first route-level characteristic is an indicator variable that is equal to one when a non-Southwest low-cost carrier is present on the route and zero otherwise. The second relevant route-level characteristic is a measure of the route Herfindahl index by quarter, where the calculated Herfindahl index is simply the sum of the squared route shares all for carriers on the route. Finally, I create variable that indicates the presence of Southwest Airlines on the route. This variable is necessary in the fixed-effects analysis to allow for accurate estimation of the correlation between the presence of non-Southwest low-cost carriers, which will allow for accurate estimation of the presence of such on the magnitude of the Southwest Effect.

Table 2.1 shows the summary statistics for all domestic routes in the United States between

[3]Southwest flights began on December 13, 2005 on routes affected by the first part of the repeal and on October 19, 2006 on routes affected by the second part of the repeal. In the strictest sense, the fourth quarters of 2005 and 2006 were quarters that had both pre-repeal and post-repeal tickets. However, the vast majority of these flights would have occurred under just one of the regimes, so the quarter is defined using the definition applicable to the majority of the tickets in that quarter.

[4]This ticket, while used to calculate route share in each of the Tampa-to-Charlotte, Charlotte-to-Chicago, and Chicago-to-New York routes, would not be included as an observation for an analysis of fares in all three routes. It would instead be an observation only for the Tampa-to-New York route since Tampa and New York are the listed origin and destination, respectively.

the first quarter of 1999 and the fourth quarter of 2009, where the unit of analysis is an airline on a route in a specific quarter-year. The first half of the table displays variables from the time period before the repeal, and the second half displays variables from the time period after the repeal. From Table 2.1, we can see that there exists substantial variation in the market characteristics of domestic routes. The route Herfindahl index ranges is around 0.7 on average, but has a standard deviation of roughly 0.25. The route share variable has an average of roughly 0.5 with a standard deviation of 0.35. Average airport share is lower, and has an average of about 0.2 with a standard deviation of roughly 0.17. There is a large change in the percentage of passengers flying on a low-cost carrier in the pre-repeal and post-repeal time periods—from the first quarter of 1999 to the third quarter of 2006, the average passenger share of low-cost airlines rises from 18% to 29%. Finally, approximately 1% of my post-repeal sample consists of observations on a DFW route affected by entry due to the repeal of the Wright Amendment.

2.3 Empirical Methods and Results

The empirical analysis has two parts. First, I estimate the impact of entry by Southwest in a "nearby" route on fares in a total of 54 routes.[5] I do so by treating each route as a separate case study and using a difference-in-difference analysis to estimate the causal effect of Southwest's entry on fares in nearby routes. The results of these analyses show that the estimated Southwest Effect differs substantially across routes. I then perform a fixed-effects analysis to uncover the market characteristics that influence the size of this effect.

2.3.1 Case Study Methods

I use a difference-in-difference approach to estimate the size of the Southwest Effect on each of the 54 affected routes. The general idea behind the difference-in-difference approach is the following. Suppose that we want to estimate the impact of some event on airfares in a route. If we look across time, there are two effects that we might see: a time effect and an event effect. The time effect is simply what would have happened had no event occurred — fares might have risen as a result of an increase in fuel prices, for example. The event effect is the effect that we really care about: the change in airfares that directly resulted from the event. Simply looking at mean airfares

[5]Recall that DFW to Phoenix, for example, is considered a different route than Phoenix to DFW.

in a route is going to be a problematic estimate of the event effect precisely because it will also be contaminated with this time effect. As such, the difference in means will only give us the correct event effect when the time effect is zero.

The problem that arises, then, is that we often have no reason to believe that the time effect truly is zero. This is where the second difference in the difference-in-difference approach comes into play: we can use routes that we think are likely to have the same time effect to estimate the time effect for the route in question. That is, if we think that other routes are unaffected by the event and behave in the same way that the affected route would have in the absence of the event, then we can estimate the time effect by looking at the difference in means for these routes (the "control" routes). If this difference accurately estimates the time effect for the route of interest, we can subtract it from the difference in means of that route to isolate the event effect.

As the choice of appropriate controls is critical to the validity of the difference-in-difference approach, I impose two criteria on possible controls. First, I stipulate that a control route must have George Bush Intercontinental Airport (IAH) in Houston, Texas as one of its endpoints. This criterion is imposed for two reasons. First, DFW and IAH are geographically close (around 4 hours apart), and so will likely suffer from the same region-specific shocks. Second, both IAH and DFW are served by a dominant carrier.[6] The second imposed criterion is a historical similarity between the changes in average fares in an affected route and its control routes. To find the routes that most meet this criterion, I calculate in both the 54 affected routes and all possible control routes the quarterly change in log average fares from the first quarter of 1999 to the fourth quarter of 2004. Formally, for each of the DFW routes and all possible controls, I calculate the squared quarterly fare change difference (SQFCD) as

$$\text{SQFCD} = \sum_{Q=1999,Q1}^{2004,Q4} [\Delta ln(avgfare_{Q+1}(DFW)) - \Delta ln(avgfare_{Q+1}(IAH))]^2, \tag{2.1}$$

where Q is the quarter between the first quarter of 1999 and the last quarter of 2004, $\Delta ln(avgfare_{Q+1}(DFW))$ is the change in log average fare in a specific DFW route between quarters $Q+1$ and Q, and $\Delta ln(avgfare_{Q+1}(IAH))$ is the change in log average fare in a specific potential control route route between quarters $Q+1$ and Q. The three George Bush Intercontinental routes with the lowest SQFCD values for a particular DFW route are assigned as the control routes for

[6]In my sample, Continental had a 62% market share at IAH in the first quarter of 2005 and Delta has a 71% market share in DFW for the same time period.

that DFW route.

Once the controls are chosen, I have 54 samples, each consisting of an affected route and its three controls. In each sample, I formally estimate the Southwest Effect in that route using the following difference-in-difference specification:

$$\begin{aligned} ln(fare_i) &= \alpha_0 + \alpha_1 * 1[\text{Post Repeal}]_i * 1[\text{DFW route}]_i \\ &+ \alpha_2 * 1[\text{DFW route}]_i + \sum_{r=1}^{2} \beta_r * 1[\text{Control r}_i]_i \\ &+ \sum_{k=1}^{43} \gamma_k * 1[\text{Quarter year k}]_i + \delta * \mathbf{X_i} + \epsilon_\mathbf{i}, \end{aligned} \tag{2.2}$$

where i denotes an individual ticket on the route, 1[Post Repeal] indicates a post-repeal quarter, 1[DFW route=1] indicates a DFW route, 1[Control r_i] indicates a ticket on control route r (one of the control routes is omitted), 1[Quarter year k] indicates that the ticket was purchased for a flight flown in quarter year k, where the second quarter of 1999 is quarter 1, the fourth quarter of 2009 is quarter 43, and the first quarter of 1999 is omitted. $\mathbf{X_i}$ is a set of other control variables consisting of whether the ticket was a roundtrip ticket and a set of airline-specific fare class indicator variables. The coefficient of interest is α_1, which reports the estimated Southwest Effect. α_2, β_1, and β_2 identify the difference in average fare levels of the DFW route and two of the control routes (respectively) relative to the omitted control route. γ_k is the average difference in levels of fares in quarter k relative to the (omitted) first quarter of 1999.

The coefficients on these sets of route and time indicator variables absorb differences in mean fare levels as well as common shocks in specific time periods, neither of which should affect the percentage change in fares resulting from entry by Southwest. The airline-specific fare class indicator variables are included to account for any changes in fares resulting from a change in the quality of tickets sold.[7] Finally, since individual tickets are likely to be correlated within a specific route in a quarter (some of the tickets likely belong to passengers on the same flight), the standard errors are clustered on the route-year-quarter level.

[7]While this would be an interesting phenomenon, study of non-fare responses to the introduction of Southwest as a competitor is outside the scope of this paper.

2.3.2 Estimating the Size of the Southwest Effect on Individual Routes

The results from the estimation of equation (2.3) on each of the 54 affected routes are shown in Table 2.2. Since the dual purposes of this part of the analysis are to give the reader a sense of the general magnitude of the Southwest Effect and to illustrate the degree of dispersion in the size of the effect across routes, only the estimated coefficients reflecting this effect are reported.

The estimated Southwest Effect is fairly large, with a median estimated percentage change of -11.75%. The Missouri routes have among the largest fare responses (fare reductions between 26% and 41%), which is perhaps unsurprising given that the ex-ante predicted gains from Southwest's entry into the competing route had to be large enough to convince the Missouri Congressman to expend the political capital necessary to achieve the partial repeal. The most notable feature of this table is the wide range of predicted effects, from a fall in fares of roughly 41% to a slight (statistically insignificant) rise in fares of 2.5%. This range in predicted effects is similar to the range found by Dresner et al. (1996) in his comparison of pre- and post-entry means. These should give pause to researchers attempting to use a small number of case studies to make more general statements about the impact of entry by Southwest—with a small number of routes, one could unknowingly get estimates that are at just one end of what is estimated here to be a wide range of effects.

A sample of full estimates is shown in Table 2.3. This table displays results for routes for which DFW is the destination, and is equivalent to the first ten estimates shown in the appropriate column of Table 2.2. We see that fares for the DFW routes are typically higher than the fares of the omitted control route. Further, fares on non-DFW routes generally do not change in a statistically significant way before and after the time period surrounding the repeal. Finally, the one-way portions of roundtrip tickets are generally substantially cheaper than an equivalent one-way only ticket.

The handful of precisely estimated non-responses in fares is particularly surprising given the presumed bias arising from route entry selection on the part of Southwest—we would expect that Southwest would choose to enter only the most profitable routes. Further, since planes are inherently mobile, the planes being used on these routes were probably employed on non-Love Field routes before the repeal. That I estimate Southwest Effects of essentially zero suggests that real factors are causing variations in this effect even in the routes that Southwest finds most profitable to enter. These results hint at the need for a deeper analysis of these factors, which I undertake in the next section.

2.3.3 Fixed-Effects Methods

The results in the previous section show substantial variation in the estimated Southwest Effect across routes. The purpose of this section is to explain some of the causes of these variations. An understanding of the drivers behind the fare response to entry by Southwest Airlines is a potentially important factor in policy decisions. As the 2005 partial repeal of the Wright Amendment shows, policymakers are willing to pass legislation with the goal of introducing competition from Southwest. This part of the analysis will answer important questions such as whether competition from Southwest still cause a reduction in fares if carriers already compete with a different low-cost carrier.

The factors that I introduce as potentially relevant to the magnitude of the route-level Southwest Effect are time-varying market-structure characteristics recognized by the literature as being relevant to the *levels* of average fares on routes (e.g., Borenstein, 1989; Evans and Kessides, 1994; Borenstein and Rose, 1994). These three factors are the route share of the carrier operating the flight, the route Herfindahl index, and the average airport share of the carrier operating the flight. These three variables are intended to capture three different dimensions of competition that might affect fares. The route Herfindahl index measures the level of competition in the route. The route share variable measures the size the carrier for whose flight the ticket was sold and is thus, to some degree, a measure of airline-specific pricing power. Finally, average airport share measures the degree of airport-level market power held by the carrier in question. By themselves, we would expect more market power to be correlated with higher prices since the route is presumably less competitive. Since even under a simple Cournot framework we would expect entry to lower prices, and to lower prices by a larger percentage when fewer firms are in the market, one might expect a higher route Herfindahl index to be correlated with a larger observed "Southwest Effect," for example. Average airport share is used instead of origin share and destination share separately because of concerns about the degrees of freedom in my estimation. While the regression has a large number of observations in total, only a handful of those observations can be used to estimate the causes of different magnitudes of the Southwest Effect. As Table 2.4 shows, the standard errors of the coefficients for variables interacted with the Southwest Effect tend to be large even when a small number of interactions are used.

Another market characteristic that I consider is whether a low-cost carrier is already present in the route at the time of Southwest's entry. This is included to measure what effect, if any, Southwest has on markets that have already adjusted to competition with a low-cost carrier. Since, as stated below, I use all domestic routes in the United States for the fixed-effects analysis, I want to accurately measure the effect of low-cost carrier presence on average fares on a route so that the effect of the presence of a low-cost carrier on the Southwest Effect is accurately measured. As such, I measure the effect on average fares of the presence of Southwest as a "direct" competitor to allow for the possibility that the effect on average fares of Southwest's presence is different than that of other low-cost carriers. Finally, I also measure the Southwest Effect on fares charged by other low-cost carriers.

To estimate these effects, I conduct a fixed-effects analysis on all routes in the United States between the first quarter of 1999 and the fourth quarter of 2009. My unit of observation is the carrier-route-quarter year, and the regression is weighted by the number of passengers flying on that airline in that route during that quarter-year. The basic equation that I want to estimate is thus

$$\begin{aligned} ln(\text{Avg. fare}_{crt}) &= \alpha_0 + \alpha_1 \text{Route share}_{crt} + \alpha_2 \text{Route Herf}_{crt} \\ &+ \alpha_3 \text{Avg. Airport Share}_{crt} + \alpha_4 [PRAR]_{crt} + \beta_1 1[PRAR]_{crt} * \text{Route share}_{crt} \\ &+ \beta_2 1[PRAR]_{crt} * \text{Route Herf}_{crt} + \beta_3 1[PRAR]_{crt} * \text{Avg. Airport Share}_{crt} \\ &+ \gamma_1 \%\ \text{LCC on Route}_{crt} + \gamma_2 1[\text{Threatened Entry}]_{crt} \\ &+ \gamma_3 1[PRAR]_{crt} \%\ \text{LCC on Route}_{crt} + \gamma_4 1[PRAR]_{crt} * 1[\text{LCC}] \\ &+ e_{crt} \end{aligned} \tag{2.3}$$

where c denotes the airline, r represents the route, and t represents the time, which is measured in quarter-years. PRAR indicates a post-repeal affected route and 1[PRAR] is an indicator variable that takes a value of 1 if the route is one of the affected DFW routes after the repeal of the Wright Amendment. % LCC on Route measures the percentage of passengers on the route that are carried by low-cost carriers. 1[LCC] equals 1 if the carrier is a non-Southwest low-cost carrier. Finally, the Threatened Entry represents the time period between the first and second stages of the repeal for the DFW routes that eventually experienced entry by Southwest Airlines.

A common concern in the literature is that route share is endogenous to the fare equation, resulting in an endogenous route Herfindahl index as well, since the route Herfindahl index is a function of a carrier's route share. The reasoning behind this concern is the likely impact of lower fares on a carrier's market share. As Borenstein (1989) discusses, the ideal instrument for route share, given the data constraints faced by the majority of the literature, is one that measures the presence of competing airlines at the endpoint routes. Borenstein notes that while airport dominance effects will impact fares, these effects are not dependent on large-scale operations at *both* endpoints. The probability of entry by other firms, however, would depend on both the degree to which the observed airline has large-scale operations at both endpoints and the degree to which other airlines do as well. The route share variable proposed as an instrument by Borenstein, then, is the geometric mean of enplanements at the origin and destination of a route divided by the sum of the same for all airlines serving that route. From equation (2.2) in Borenstein (1989), This geometric enplanement share ($GENPSH$) is defined as

$$GENPSH = \frac{\sqrt{ENP_{i1} * ENP_{i2}}}{\sum_j \sqrt{ENP_{j1} * ENP_{j2}}}, \tag{2.4}$$

where i is the airline in question, j denotes each airline that services the route, and ENP_{j1} and ENP_{j2} are the number of enplanements on airline j at the origin and destination airports, respectively. Crucially, $GENPSH$ will differ from the average airport share variable included in the second stage analysis because it measures the "dual large-scale operations" effect. If the airline has a large presence in one airport but a small presence in the other, the small presence at one endpoint will affect route share simply by capping the feasible number of flights that the airline can make. Borenstein then includes another measure of potential competition in an attempt to absorb any direct effects that this instrument might have on fares. The average airport share variable in this analysis plays that role.[8]

As I am interested in the effect of route share on the magnitude of the "Southwest Effect," I estimate the effect of the interaction term $1[PRAR] *$ Route share as well. The instrument for this variable is simply the interaction between the indicator variable for a flight on a post-repeal affected route ($1[PRAR]$) and the geometric enplanement share ($GENPSH$). The first stage regressions for

[8]Borenstein admits, as I will do here, that this is not a perfect instrument. However, he (as well as the rest of the literature) feels that using the instrument is better than ignoring the almost-certain endogeneity issues. Previous instruments included such measures as the log of market share and the route Herfindahl index, but these instruments are problematic as well because of the possibility of brand loyalty (induced by frequent flier programs, for example) resulting in a causal effect of past market share and Herfindahl index level on current fares.

route share and its interaction with Southwest's entry are then

$$\begin{aligned} \text{Route share}_{crt} &= \alpha_1 + \beta_{11} GENPSH_{crt} + \beta_{12} 1[PRAR] * GENPSH_{crt} \\ &+ \gamma_1 \mathbf{X}_{crt} + \epsilon_{1,crt} \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} 1[PRAR] * \text{Route share}_{crt} &= \alpha_2 + \beta_{21} GENPSH_{crt} + \beta_{22} 1[PRAR] * GENPSH_{crt} \\ &+ \gamma_2 \mathbf{X}_{crt} + \epsilon_{2,crt}, \end{aligned} \tag{2.6}$$

where again c denotes the airline, r represents the route, t represents the time period, and $\mathbf{X}$ denotes the exogenous second-stage variables.

Once the predicted values of route share have been obtained, the instrument for the route Herfindahl index can be generated. Borenstein suggests a "rescaled" route Herfindahl measure that takes the predicted route share of the observed airline as fixed and then "rescales" the shares of the other carriers such that the measure once again ranges from zero to one. From equation (2.3) in Borenstein (1989), the instrument for the route Herfindahl index (IRUTHERF) is formally generated as

$$IRUTHERF = \widehat{\text{Route share}}^2 + \frac{\text{Route Herf - Route Share}^2}{(1 - \text{Route Share})^2} * (1 - \widehat{RouteShare})^2. \tag{2.7}$$

Here, $\widehat{\text{Route Share}}$ is the predicted route share from the first stage regression. As Borenstein discusses, this instrument is valid if geometric enplanement share instrument is valid for route share and if the price of an airline's tickets does not affect the way in which consumers purchasing tickets from rival airlines sort between these rivals.

Since, as with market share, I am interested in the effect of the route Herfindahl index on the magnitude of the "Southwest Effect," I generate an instrument for the post-repeal Herfindahl index on affected routes ($1[PRAR] *$ Route Herf) by interacting $IRUTHERF$ and the post-repeal affected route indicator variable. The first-stage regressions for the route Herfindahl index and its interaction with the DFW routes affected by competition from Southwest are then

$$\begin{aligned}\text{Route Herf}_{crt} &= \alpha_3 + \beta_{31} IRUTHERF_{crt} + \beta_{32} 1[\text{PRAR}] * IRUTHERF_{crt} \\ &+ \gamma_3 \mathbf{X}_{crt} + \epsilon_{3,crt}\end{aligned} \tag{2.8}$$

and

$$\begin{aligned}1[\text{PRAR}] * \text{Route Herf}_{crt} &= \alpha_4 + \beta_{41} IRUTHERF_{crt} \\ &+ \beta_{42} 1[\text{PRAR}] * IRUTHERF_{crt} + \gamma_4 \mathbf{X}_{crt} + \epsilon_{4,crt},\end{aligned} \tag{2.9}$$

where again c denotes the airline, r represents the route, t represents the time period, and $\mathbf{X}$ denotes the exogenous second-stage variables.

Using the first-stage regressions in equations (2.5)-(2.9) to generate predicted values of the endogenous variables and then inserting these predicted values into equation (2.3) yields the second-stage regression that I estimate:

$$\begin{aligned}ln(\text{Avg. fare}_{crt}) &= \alpha_0 + \alpha_1 \widehat{\text{Route share}}_{crt} + \alpha_2 \widehat{\text{Route Herf}}_{crt} \\ &+ \alpha_3 \text{Avg. Airport Share}_{crt} + \beta_1 1[\text{PRAR}]_{crt} \widehat{* \text{Route}} \text{ share}_{crt} \\ &+ \beta_2 1[\text{PRAR}]_{crt} \widehat{* \text{Route}} \text{ Herf}_{crt} + \beta_3 1[\text{PRAR}]_{crt} * \widehat{\text{Avg. }} \text{Airport Share}_{crt} \\ &+ \gamma_1 1[\text{LCC on Route}]_{crt} + \gamma_2 1[\text{Southwest on Route}]_{crt} \\ &+ \gamma_3 1[\text{PRAR}]_{crt} * 1[\text{LCC on Route}]_{crt} + \gamma_4 1[\text{PRAR}]_{crt} * 1[\text{LCC}] \\ &+ e_{crt}\end{aligned} \tag{2.10}$$

where c denotes the airline, r represents the route, and t represents the time, which is measured in quarter-years. Several sub-specifications of this equation are estimated, both for robustness and to conserve degrees of freedom. While the database has a large number of observations, there are only a few hundred observations that occur on the routes that are directly affected by the Wright Amendment.

2.3.4 The Effect of Market Structure on the Southwest Effect

Tables 2.4 and 2.5 shows the results of the estimation of equation (2.10) with and without the “Threatened Entry” variable. As the results of the two tables are similar, I will focus on Table 2.4. Four specifications are shown. The first, in column (1), looks only at the estimated Southwest Effect as well as the impact on fares of the route Herfindahl index, route share, and average airport share. Fares decrease by roughly 25% due to increased competition from Southwest Airlines in a nearby route. This number is very similar to the 26.4% effect predicted by Morrison (2001) on this form of “nearby” competition. The route Herfindahl index and average airport share of the airline have large positive estimated effects on the levels of fares in routes, meaning that lower observed levels of competition are correlated with higher average fares. The estimated effect of route share is actually *negative*, which is curious but not unheard-of in the literature—Evans and Kessides (1994) find a small negative effect for this variable after accounting for average route share as well. Routes with a higher share of passengers flying low-cost carriers tend to have lower fares — this sign is expected since the % LCC on Route variable likely serves as a proxy for the extent to which airlines must compete with low-cost carriers. Finally, the Threatened Enrty variable indicates that fares fell in the affected routes by about 13.5% in the period between the first and second stages of the repeal of the Wright Amendment. This magnitude is consistent across specifications, and contradicts the predictions of Goolsbee and Syverson (2008), which would have suggested that “threatened” entry would have no effect.

Column (2) of Table 2.4 uses the explanatory variables from column (1) and adds variables indicating whether there is a (non-Southwest) low-cost carrier on the route, whether Southwest services the route directly, and two variables that look at the impact of the presence of low-cost carriers on the Southwest Effect.[9] The first, the size of existing low-cost carriers on the affected DFW routes, is positive and statistically significant. This estimate shows that changing the percentage of passengers flying on low-cost carriers from 0% to 10% will reduce the estimated Southwest Effect by about 5 percentage points. This dampening of the Southwest Effect is likely due to the lower fares resulting from competition with the existing low-cost airlines. Further, the if a ticket is purchased on a low-cost airline, the effect is dampened by another roughly four and a half percentage points,

[9] While I am now estimating the correlation between fares and Southwest’s presence on routes directly, I will still refer to the “Post-repeal on an affected route” coefficient as the Southwest Effect since I am still using the Wright Amendment as a natural experiment to try to get at causality, whereas the correlation between fares and the presence of Southwest on a route is unlikely to be a causal relationship due to a host of other factors that affect the entry and exit decisions of both Southwest and traditional airlines.

although this effect is not statistically significant.

Column (3) of Table 2.4 uses the explanatory variables from column (1) as well as estimating the effects of route Herfindahl index, route share, and average airport share on the magnitude of the Southwest Effect. Route Herfindahl index has an effect on the fare response that is statistically significant but very small—changing the route from perfectly competitive to monopolistic will only increase the size of the Southwest Effect by about one percentage point, holding route share and average airport share constant. Since route share causes the effect to be slightly dampened, if we change route share from 0 to 1 as we change route Herfindahl index from 0 to 1 (which would be necessary to in a market served by a monopolist), the net effect on the magnitude of the Southwest Effect is a roughly 3.5 percentage point mitigation of the effect, although this is not statistically significant. Average airport share, on the other hand, is both statistically economically significant. Changing the average airport share from 0 to 1 increases the magnitude of the Southwest Effect by around 32 percentage points. So, this specification shows us that route Herfindahl index and route share have negligible effects on the magnitude of the "Southwest Effect," but average airport share is a very important determinant.

Estimation of the full specification shown in equation (2.10) is reported in column (4) of Table 2.4. In this specification, the factors that were shown in column (3) to have a relatively insignificant effect on the size of the Southwest Effect are still estimated to have a relatively insignificant effect. The presence of an existing low-cost carrier retains its estimated impact. The effect of average share on the size of the "Southwest Effect," however, falls by almost half and loses statistical significance. The point estimate, however, still retains economic significance—a change in average airport share from 0 to 1 increases the size of the estimated Southwest Effect by about 16 percentage points.

These results, then, point to two important determinants of the size of the Southwest Effect: the presence of a low-cost carrier and the average airport share of the airline servicing the ticket. Intriguingly, two other factors (route share and route Herfindahl) that are generally found to impact the *level* of fares are shown to have almost almost zero impact on the *change* in fares due to entry by Southwest on a "nearby" route.

2.4 Conclusion

This paper fills a gap in the literature by illustrating the dispersion in the estimated Southwest Effect across routes and by analyzing the factors that lead to this dispersion. Using a natural experiment created by the repeal of the Wright Amendment in Dallas, Texas, I find that two factors are especially important to the magnitude of the Southwest Effect. The first, the presence of a low-cost carrier on a route, tends to dampen the effect by about 10 percentage points. The second, average airport share, tends to magnify the effect by 18 to 32 percentage points for a 0 to 1 increase in average airport share (that is, essentially by moving from an airline that enplanes only one passenger in the two endpoint airports to one that flies all of the passengers in the two endpoint airports). As the observed range of average airport market share is 0 (rounded to two decimal places) to 1, this factor has the potential to account for large differences in the observed Southwest Effect. I also find that route Herfindahl index and route share, two important factors in determining the *average* fare on a route, seem to have little effect on the *change* in fares on a route due to the introduction of Southwest as a competitor in a "nearby" route.

Since the first part of the repeal of the Wright Amendment was the result of legislative action to allow Southwest Airlines to service airports in Missouri from Love Field, a better understanding of the determinants of the magnitude of the Southwest Effect is relevant to policy makers. My results have two important policy implications. First, even though the presence of low-cost carriers mitigates the gains from competition from Southwest Airlines, passengers on routes with existing low-cost carriers are still likely to benefit significantly from competition from Southwest Airlines. Second, I find that the average airport share of the airline operating the ticket has a large impact on the change in fares charged by that airline. Thus, legislation encouraging entry into routes operated by airlines with high average airport shares is likely to lead to the largest reduction in fares.

2.5 Tables and Figures

Table 2.1: Summary Statistics

Variable	Mean	Std. Dev.	Mean	Std. Dev.
	Pre-Repeal		Post-Repeal	
Log Fare	5.184	0.376	5.214	0.337
Route Share	0.530	0.396	0.536	0.388
Route Herfindahl	0.719	0.244	0.686	0.264
Avg. Airport Share	0.212	0.162	0.213	0.175
% LCC Tickets on Route	0.183	0.291	0.290	0.335
Affected Route			0.012	0.107
N	308820		132349	

The level of observations is a carrier-route-quarter-year. The reported statistics are weighted by the number of passengers.

Table 2.2: Route-Specific Estimates of the Southwest Effect

	DFW as origin airport			DFW as destination airport		
Endpoint	DD estimate	S.E.	% Change	DD estimate	S.E.	% Change
St. Louis	−0.446***	0.039	−35.98%	−0.525***	0.035	−40.84%
Kansas City	−0.302***	0.048	−26.07%	−0.379***	0.042	−31.55%
Cleveland	−0.064	0.044	−6.20%	−0.235***	0.037	−20.94%
Columbus	−0.091*	0.035	−8.70%	−0.120**	0.039	−11.31%
Denver	−0.264***	0.035	−23.20%	−0.119***	0.024	−11.22%
Detroit	−0.105*	0.048	−9.97%	−0.264***	0.032	−23.20%
Ft. Lauderdale	−0.087*	0.039	−8.33%	−0.002	0.026	−0.20%
Indianapolis	−0.065	0.054	−6.29%	−0.208***	0.032	−18.78%
Jacksonville	−0.064**	0.022	−6.20%	0.025	0.030	2.53%
Las Vegas	−0.011	0.057	−1.09%	−0.007	0.029	−0.70%
Los Angeles	0.022	0.039	2.22%	0.010	0.029	1.01%
Louisville	−0.104**	0.036	−9.88%	−0.228***	0.036	−20.39%
Nashville	−0.356***	0.022	−29.95%	−0.311***	0.032	−26.73%
Oakland	−0.232***	0.036	−20.71%	−0.182***	0.037	−16.64%
Omaha	−0.196***	0.033	−17.80%	−0.229***	0.032	−20.47%
Orlando	−0.014	0.035	−1.39%	−0.005	0.023	−0.50%
Philadelphia	−0.103*	0.041	−9.79%	−0.163***	0.031	−15.04%
Phoenix	−0.169***	0.031	−15.55%	−0.124***	0.029	−11.66%
Portland	−0.038	0.038	−3.73%	−0.065*	0.026	−6.29%
Sacramento	−0.097**	0.035	−9.24%	−0.043	0.028	−4.21%
Salt Lake City	−0.239***	0.024	−21.26%	−0.132***	0.024	−12.37%
San Diego	−0.117*	0.048	−11.04%	−0.118***	0.026	−11.13%
Seattle	−0.079*	0.036	−7.60%	−0.125***	0.028	−11.75%
Tampa Bay	−0.133***	0.033	−12.45%	−0.035	0.027	−3.44%
Tuscon	−0.122***	0.036	−11.49%	−0.127***	0.036	−11.93%
Washington	−0.079*	0.037	−7.60%	−0.064	0.034	−6.20%
Chicago	−0.302***	0.035	−26.07%	−0.233***	0.042	−20.78%

Standard errors are clustered by route-quarter-year. ***, **, and * represent statistical significance at the .1%, 1%, and 5% levels, respectively. The dependent variable is log ticket fare. The reported coefficients are estimates of the effect on fares of entry by Southwest in a "nearby" route. All regressions also include route-specific and quarter-specific indicator variables, as well as variables indicating whether the ticket is a roundtrip ticket and the airline-specific fare class of the ticket.

Table 2.3: Sample of Complete Difference-in-Difference Estimates with DFW as Destination Airport

	St. Louis	Kansas City	Cleveland	Columbus	Denver	Detroit	Ft. Lauderdale	Indianapolis	Jacksonville	Las Vegas
Post*DFW	-0.525^{***}	-0.379^{***}	-0.235^{***}	-0.120^{**}	-0.119^{***}	-0.264^{***}	-0.002	-0.208^{***}	0.025	-0.007
	(0.035)	(0.042)	(0.037)	(0.039)	(0.024)	(0.032)	(0.026)	(0.032)	(0.030)	(0.029)
DFW	-0.066^{**}	0.077^{*}	0.139^{***}	-0.033	0.056^{**}	0.231^{***}	0.124^{***}	-0.052^{*}	0.098^{***}	0.160^{***}
	(0.023)	(0.030)	(0.017)	(0.031)	(0.018)	(0.017)	(0.017)	(0.022)	(0.024)	(0.025)
Post	0.03	-0.024	-0.036	-0.062	-0.105^{*}	-0.015	-0.031	0.043	-0.084	-0.063
	(0.045)	(0.041)	(0.097)	(0.053)	(0.044)	(0.044)	(0.034)	(0.047)	(0.047)	(0.033)
Roundtrip	-0.278^{***}	-0.277^{***}	-0.411^{***}	-0.337^{***}	-0.279^{***}	-0.267^{***}	-0.328^{***}	-0.304^{***}	-0.360^{***}	-0.293^{***}
	(0.013)	(0.008)	(0.014)	(0.010)	(0.008)	(0.009)	(0.010)	(0.013)	(0.010)	(0.012)
Constant	6.314^{***}	5.446^{***}	7.364^{***}	5.949^{***}	5.871^{***}	6.242^{***}	5.853^{***}	5.970^{***}	6.124^{***}	6.234^{***}
	(0.041)	(0.161)	(0.025)	(0.670)	(0.036)	(0.517)	(0.172)	(0.161)	(0.143)	(0.148)
R^2	0.367	0.257	0.300	0.266	0.248	0.300	0.254	0.267	0.252	0.290
N	212866	401268	295397	252436	472868	213997	462442	168694	407847	583711

Standard errors in parentheses. ***, **, and * represent statistical significance at the .1%, 1%, and 5% levels, respectively. The dependent variable is log ticket fare. The reported coefficients are estimates of the effect on fares of entry by Southwest in a “nearby” route. Standard errors are clustered by route-quarter-year.

Table 2.4: The Impact of Market Characteristics on the Size of the Southwest Effect

	(1)	(2)	(3)	(4)
Post-Repeal Affected Route (PRAR)	-0.293^{***}	-0.313^{***}	-0.216^{***}	-0.238^{***}
	(0.011)	(0.011)	(0.030)	(0.033)
Route Herf	0.274^{***}	0.286^{***}	0.306^{***}	0.307^{***}
	(0.018)	(0.015)	(0.016)	(0.016)
Route Share	-0.125^{***}	-0.141^{***}	-0.125^{***}	-0.126^{***}
	(0.017)	(0.016)	(0.016)	(0.016)
Avg. Airport Share	0.474^{***}	0.476^{***}	0.450^{***}	0.450^{***}
	(0.021)	(0.019)	(0.020)	(0.020)
% LCC on route	-0.230^{***}	-0.211^{***}	-0.207^{***}	-0.208^{***}
	(0.009)	(0.008)	(0.008)	(0.008)
Threatened Entry	-0.137^{***}	-0.134^{***}	-0.137^{***}	-0.135^{***}
	(0.017)	(0.015)	(0.015)	(0.015)
% LCC on route∗PRAR		0.519^{**}		0.546^{**}
		(0.136)		(0.147)
LCC airline∗PRAR		0.044		−0.014
		(0.065)		(0.074)
Route Herf∗PRAR			-0.016^{**}	-0.017^{**}
			(0.004)	(0.004)
Route Share∗PRAR			0.036	−0.018
			(0.070)	(0.073)
Avg. Airport Share∗PRAR			-0.273^{**}	−0.158
			(0.139)	(0.141)
N	441169	441169	441169	441169
R^2	0.208	0.209	0.209	0.210

Standard errors in parentheses. ***, **, and * represent statistical significance at the .1%, 1%, and 5% levels, respectively. The dependent variable is log average fare. PRAR indicates observations on Post-Repeal Affected Routes, which are the 54 DFW routes that experienced immediate competition from Southwest Airlines. Observations are weighted by number of passengers.

Table 2.5: The Impact of Market Characteristics on the Size of the Southwest Effect Without Threat

	(1)	(2)	(3)	(4)
Post-Repeal	-0.267^{***}	-0.292^{***}	-0.181^{***}	-0.206^{***}
Affected Route (PRAR)	(0.011)	(0.011)	(0.030)	(0.033)
Route Herf	0.271^{***}	0.284^{***}	0.302^{***}	0.303^{***}
	(0.018)	(0.015)	(0.016)	(0.016)
Route Share	-0.124^{***}	-0.140^{***}	-0.125^{***}	-0.126^{***}
	(0.017)	(0.016)	(0.016)	(0.016)
Avg. Airport Share	0.471^{***}	0.473^{***}	0.449^{***}	0.449^{***}
	(0.020)	(0.019)	(0.020)	(0.020)
% LCC on route	-0.229^{***}	-0.211^{***}	-0.206^{***}	-0.207^{***}
	(0.009)	(0.008)	(0.008)	(0.008)
% LCC on route∗PRAR		0.541^{**}		0.567^{**}
		(0.136)		(0.147)
LCC airline∗PRAR		0.063		−0.003
		(0.065)		(0.074)
Route Herf∗PRAR			-0.016^{**}	-0.016^{**}
			(0.004)	(0.004)
Route Share∗PRAR			0.040	−0.015
			(0.070)	(0.073)
Avg. Airport Share∗PRAR			-0.318^{**}	−0.191
			(0.139)	(0.141)
N	441169	441169	441169	441169
R^2	0.207	0.208	0.208	0.209

Standard errors in parentheses. ***, **, and * represent statistical significance at the .1%, 1%, and 5% levels, respectively. The dependent variable is log average fare. PRAR indicates observations on Post-Repeal Affected Routes, which are the 54 DFW routes that experienced immediate competition from Southwest Airlines. Observations are weighted by number of passengers.

Appendices

Appendix A

Hazard First-Stage Results for the NLSY Experience Smoothing

Figure A.1: Hazard Predictions for 40-Year-Old White College Graduates

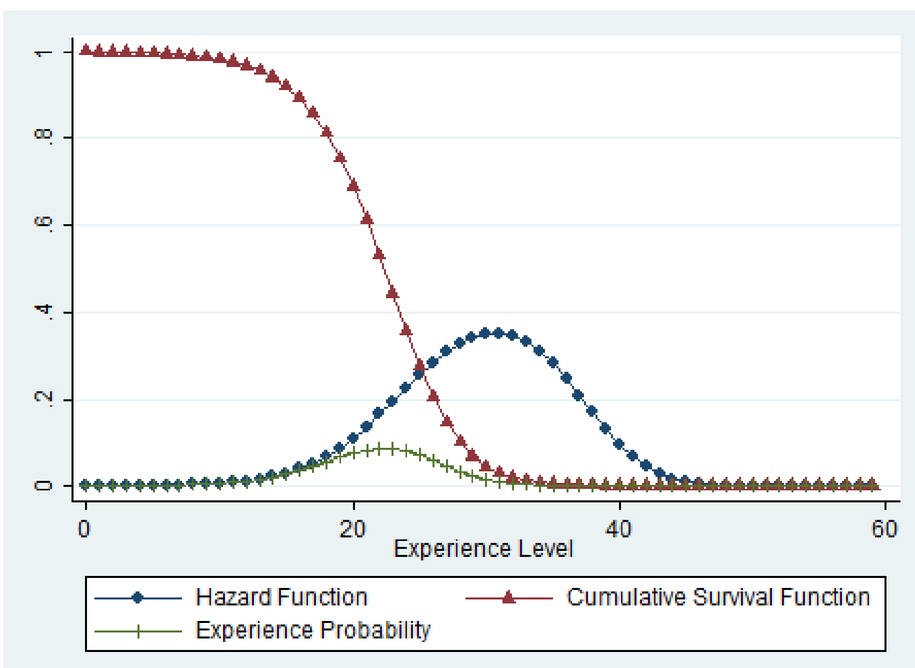

Figure A.2: Hazard Predictions for 40-Year-Old Black College Graduates

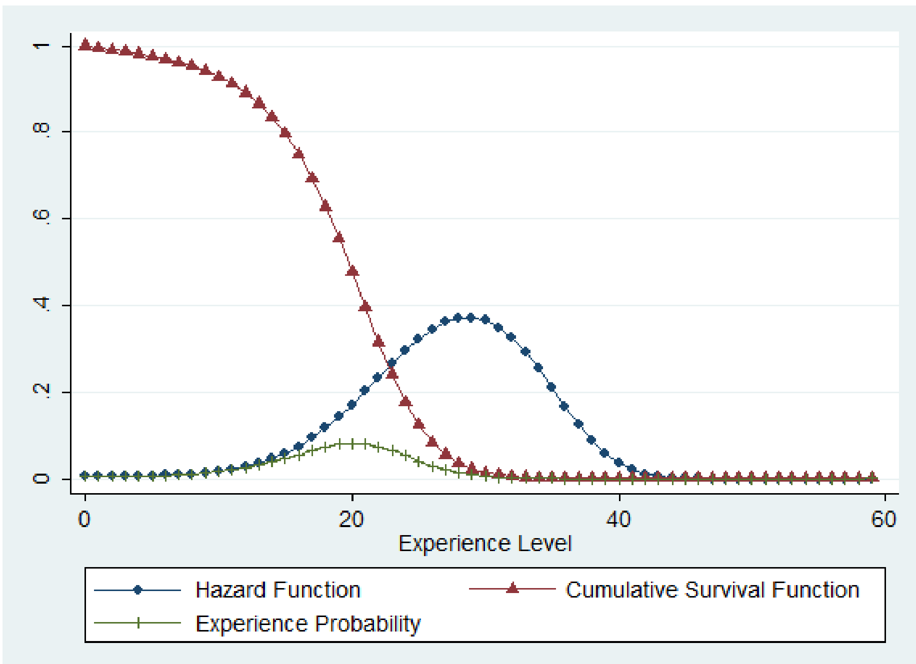

Figure A.3: Hazard Predictions for 22-Year-Old White High School Dropouts

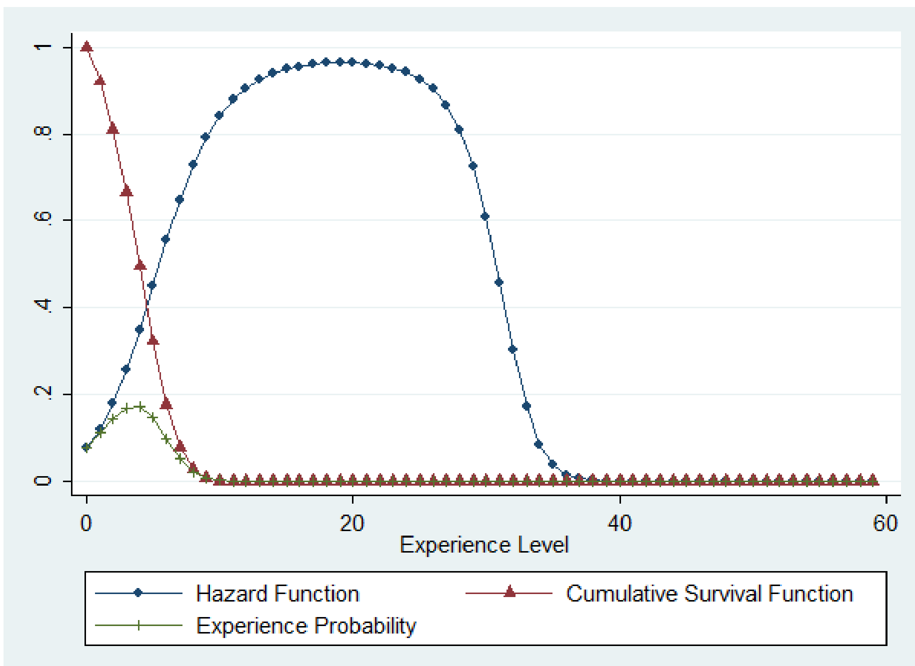

Figure A.4: Hazard Predictions for 22-Year-Old Black High School Dropouts

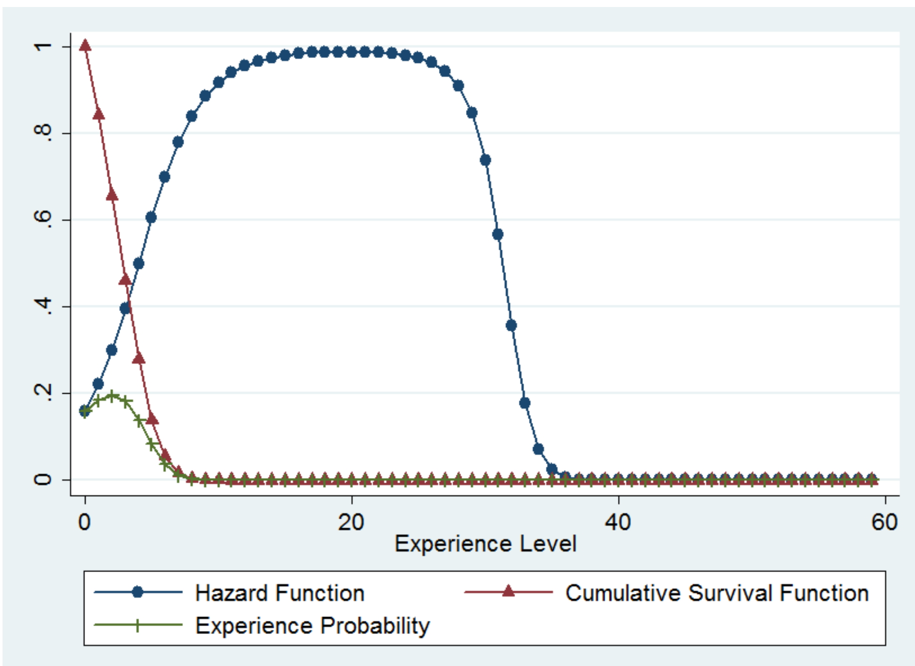

Table A.1: Hazard Regression for NLSY

	$\hat{\beta}$	$\hat{\sigma}$
ln(Age)	−674.1	(41.91)
ln(Age)^2	193.5	(12.10)
ln(Age)^3	−18.43	(1.178)
ln(Age)∗Exp	−0.152	(0.150)
ln(Age)∗Exp^2	−0.0291	(0.0120)
ln(Age)∗Exp^3	0.00106	(0.000340)
Exp	0.504	(0.509)
Exp^2	0.113	(0.0434)
Exp^3	−0.00398	(0.00127)
Black	−200.4	(71.67)
Black∗ ln(Age)	184.7	(63.23)
Black∗ ln(Age)^2	−56.93	(18.55)
Black∗ ln(Age)^3	5.899	(1.810)
Black∗Exp	0.577	(0.172)
Black∗Exp^2	0.0159	(0.0148)
Black∗Exp^3	−0.000772	(0.000427)
Black∗ ln(Age)*Exp	−0.211	(0.0506)
Black∗ ln(Age)*Exp^2	−0.00213	(0.00410)
Black∗ ln(Age)*Exp^3	0.000175	(0.000115)
Educ	2.513	(1.355)
Educ∗ ln(Age)	−0.853	(0.821)
Educ∗ ln(Age)^2	0.0236	(0.124)
Educ∗Exp	0.185	(0.0396)
Educ∗Exp^2	−0.0171	(0.00337)
Educ∗Exp^3	0.000302	(0.0000992)
Educ∗ ln(Age)∗Exp	−0.0456	(0.0116)
Educ∗ ln(Age)∗Exp^2	0.00469	(0.000930)
Educ∗ ln(Age)∗Exp^3	−0.0000853	(0.0000266)
Constant	775.1	(48.79)
N	689420	
LLF	−160098.56	

Standard errors in parentheses.

Appendix B

First-Stage Results for Multi-Sample 2SLS

Figure B.1: Predicted Experience for High School Dropouts from the NLSY Measure

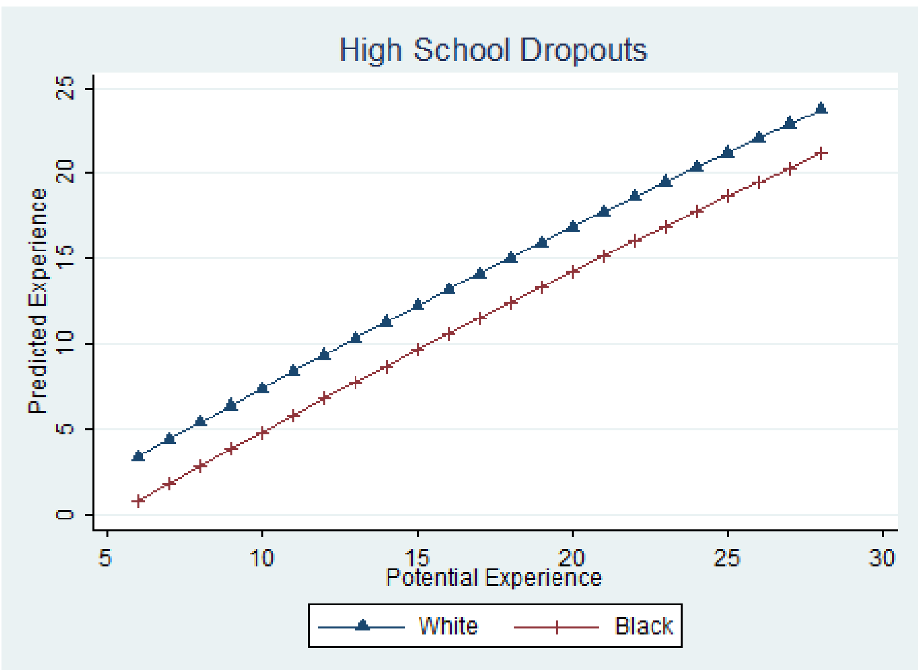

Figure B.2: Predicted Experience for High School Graduates from the NLSY Measure

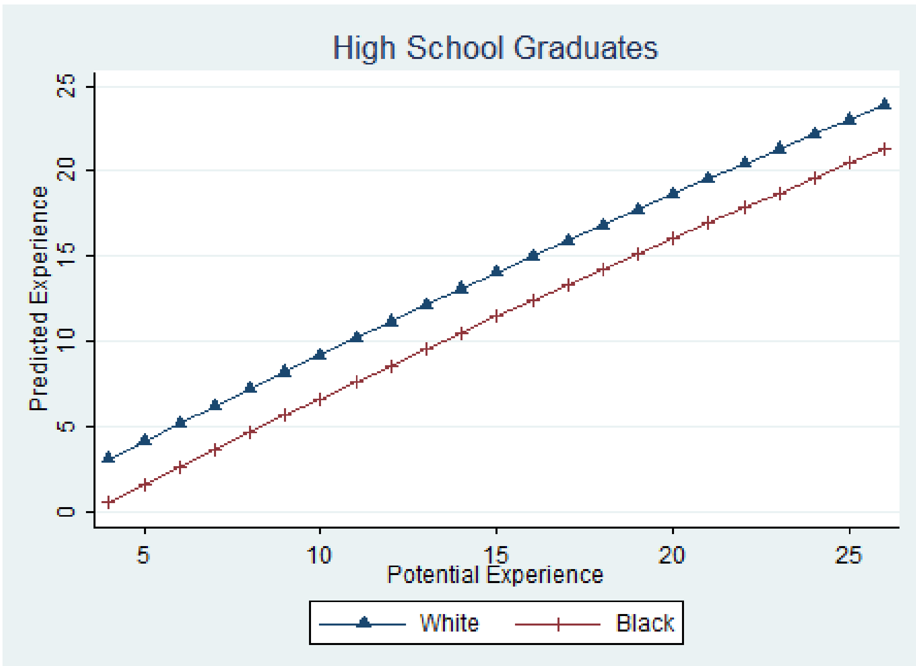

Figure B.3: Predicted Experience for Some College from the NLSY Measure

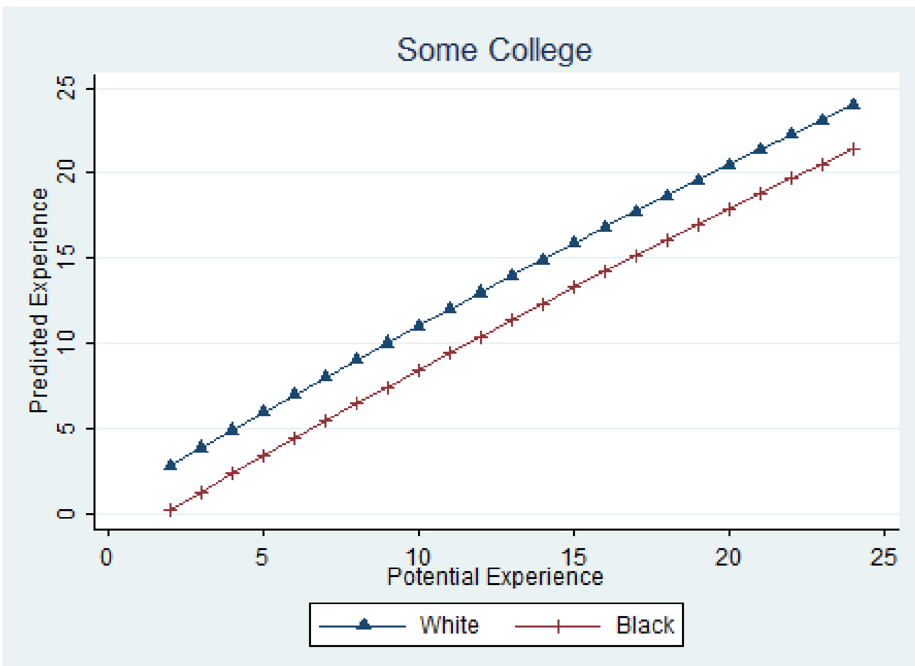

Figure B.4: Predicted Experience for College Graduates from the NLSY Measure

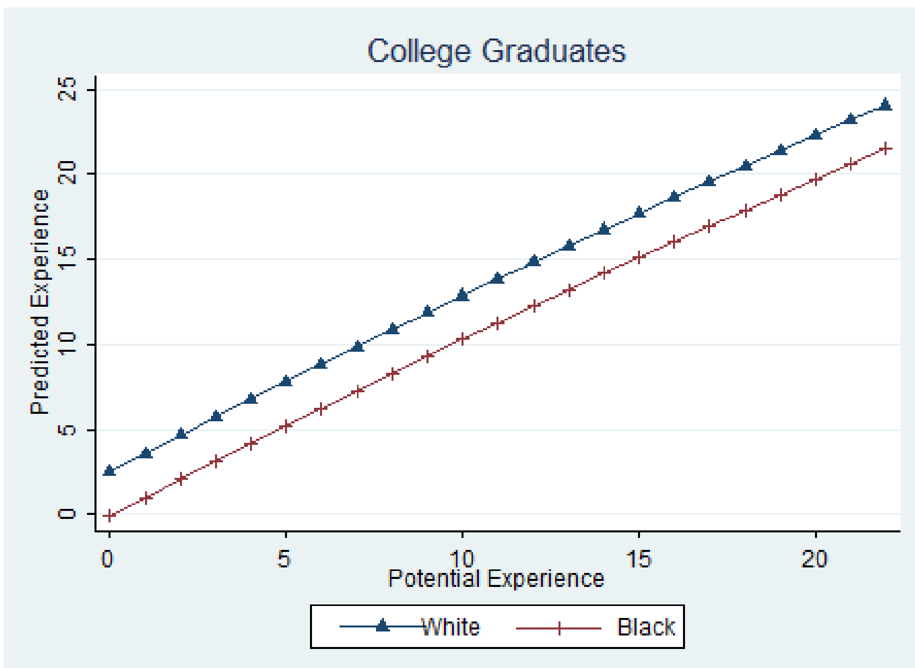

Figure B.5: Predicted Experience for High School Dropouts from the PSID Measure

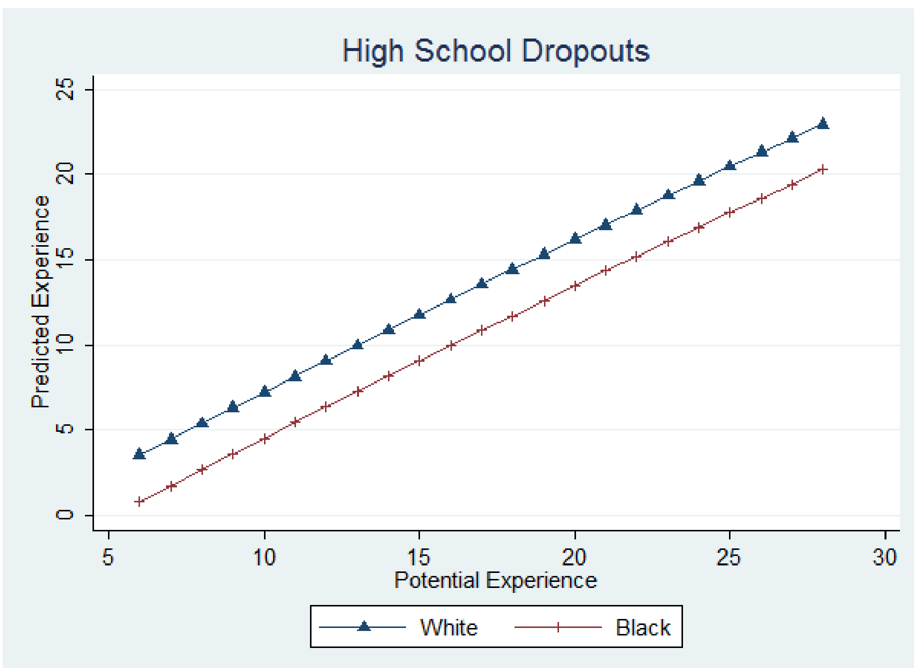

Figure B.6: Predicted Experience for High School Graduates from the PSID Measure

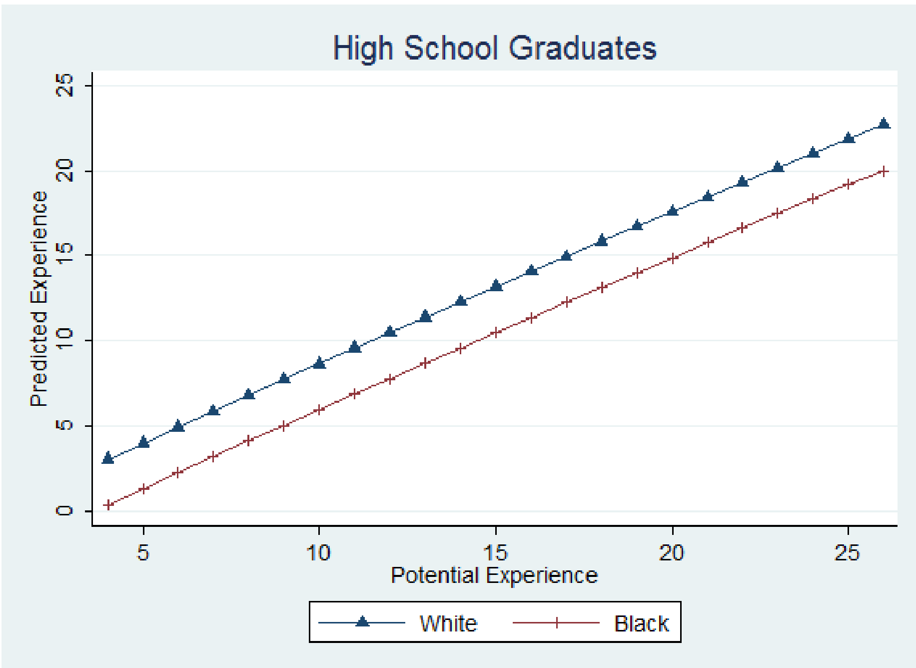

Figure B.7: Predicted Experience for Some College from the PSID Measure

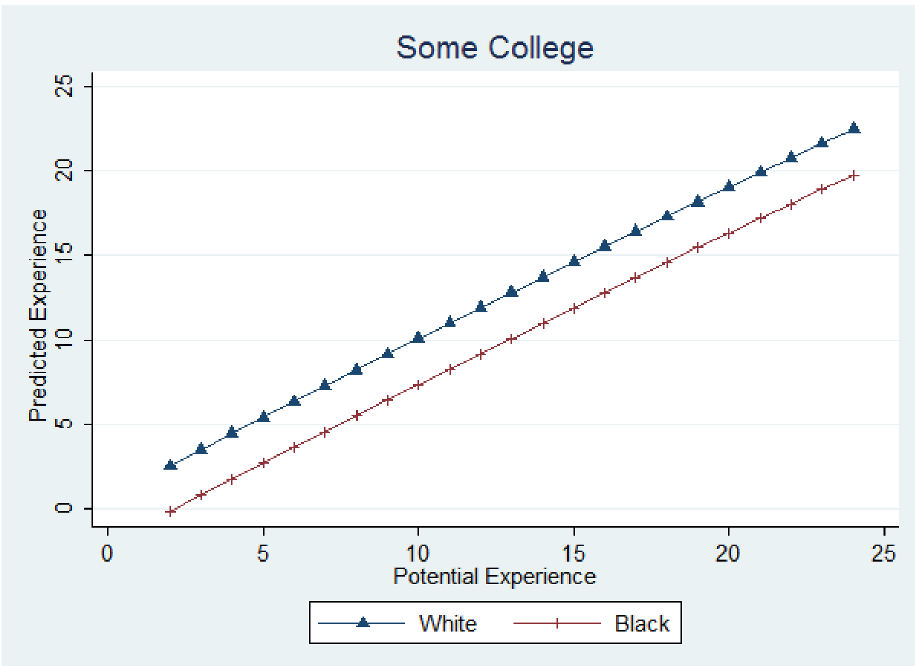

Figure B.8: Predicted Experience for College Graduates from the PSID Measure

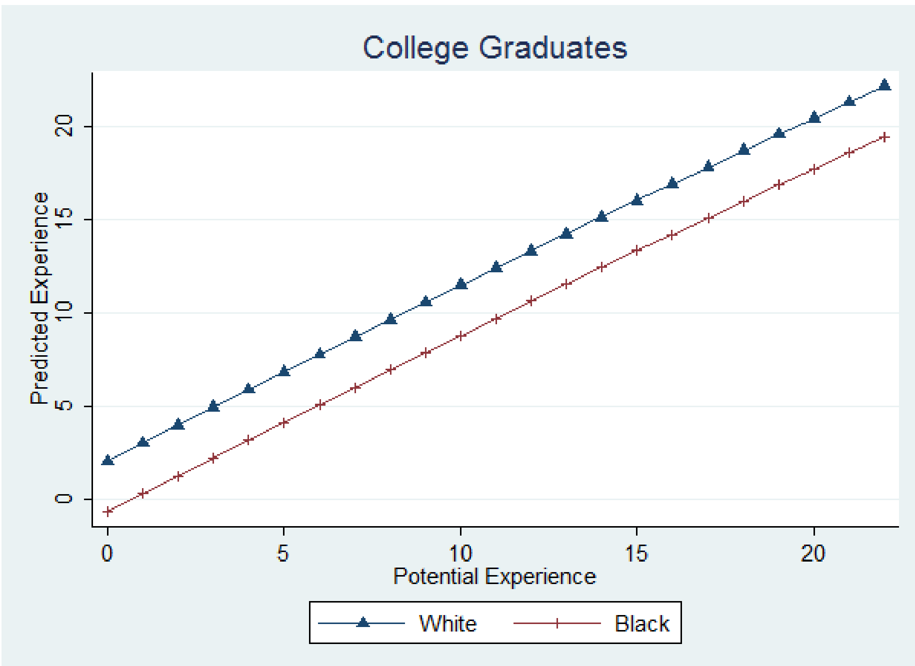

Table B.1: First-Stage Multi-Sample 2SLS Results

	Sibling Measure of Education	Enrollment Measure of Education	NLSY Experience	NLSY Experience^2	PSID Experience	PSID Experience^2
Reported Education	0.899	1.012	0.908	22.57	0.711	13.16
	(0.0249)	(0.00132)	(0.00921)	(0.300)	(0.0108)	(0.272)
Potential Experience	−0.216	−0.0116	1.081	10.27	0.969	5.317
	(0.0885)	(0.00227)	(0.0114)	(0.371)	(0.0147)	(0.372)
Potential Experience^2	0.00775	0.000671	−0.00457	0.583	−0.00252	0.608
	(0.00334)	(0.000120)	(0.000413)	(0.0134)	(0.000607)	(0.0154)
Black	−0.264	−0.0295	−2.579	−63.78	−2.706	−50.20
	(0.0721)	(0.00497)	(0.0377)	(1.228)	(0.0407)	(1.029)
Constant	2.721	−0.221	−12.03	−324.2	−9.323	−179.0
	(0.769)	(0.0222)	(0.151)	(4.913)	(0.176)	(4.462)
N	551	43283	59770	59770	28932	28932
R^2	0.838	0.943	0.654	0.569	0.691	0.620

Standard errors in parentheses.

Bibliography

Altonji, J. and R. Blank (1999). Race and gender in the labor market. *Handbook of Labor Economics 3*, 3143–3259.

Antecol, H. and K. Bedard (2004). The racial wage gap: The importance of labor force attachment differences across black, Mexican, and white men. *Journal of Human Resources 39*(2), 564.

Bennett, R. and J. Craun (1993). The airline deregulation evolution continues: The Southwest effect. *Office of Aviation Analysis, US Department of Transportation. May*.

Black, D., S. Sanders, and L. Taylor (2003). Measurement of higher education in the census and current population survey. *Journal of the American Statistical Association 98*(463), 545–554.

Blau, F. and L. Kahn (2011, July). The feasibility and importance of adding measures of actual experience to cross-sectional data collection. Working Paper 17241, National Bureau of Economic Research.

Bollinger, C. (2003). Measurement error in human capital and the black-white wage gap. *The Review of Economics and Statistics 85*(3), 578–585.

Borenstein, S. (1989). Hubs and high fares: dominance and market power in the US airline industry. *Rand Journal of Economics 20*(3), 344–365.

Borenstein, S. and N. Rose (1994). Competition and price dispersion in the US airline industry. *Journal of Political Economy 102*(4), 653–683.

Bratsberg, B. and D. Terrell (1998). Experience, tenure, and wage growth of young Black and White men. *Journal of Human Resources 33*(3), 658–682.

Cain, G. (1986). The Economic Analysis of Labor Market Discrimination: A Survey. *Handbook of Labor Economics 1*, 693–785.

Dresner, M., J. Lin, and R. Windle (1996). The impact of low-cost carriers on airport and route competition. *Journal of Transport Economics and Policy 30*(3), 309–328.

Evans, W. and I. Kessides (1994). Living by the "Golden Rule": Multimarket contact in the US airline industry. *The Quarterly Journal of Economics 109*(2), 341–366.

Filer, R. (1993). The usefulness of predicted values for prior work experience in analyzing labor market outcomes for women. *Journal of Human Resources 28*(3), 519–537.

Garvey, N. and C. Reimers (1980). Predicted vs. potential work experience in an earnings function for young women. *Research in Labor Economics 3*, 99–127.

Gilleskie, D. and T. Mroz (2004). A flexible approach for estimating the effects of covariates on health expenditures. *Journal of Health Economics 23*(2), 391–418.

Goolsbee, A. and C. Syverson (2008). How do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines. *Quarterly Journal of Economics 123*(4), 1611–1633.

Inoue, A. and G. Solon (2010). Two-sample instrumental variables estimators. *The Review of Economics and Statistics 92*(3), 557–561.

Morrison, S. (2001). Actual, Adjacent, and Potential Competition Estimating the Full Effect of Southwest Airlines. *Journal of Transport Economics and Policy 35*(2), 239–256.

Mroz, T. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica 55*(4), 765–799.

Murphy, K. and R. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics 3*(4), 370–79.

Neal, D. and W. Johnson (1996). The role of premarket factors in black-white wage differences. *The Journal of Political Economy 104*(5), 869–895.

Regan, T. and R. Oaxaca (2009). Work experience as a source of specification error in earnings models: implications for gender wage decompositions. *Journal of Population Economics 22*(2), 463–499.

Ruggles, S., J. Alexander, K. Genadek, R. Goeken, M. Schroeder, and M. Sobek (2010). Integrated public use microdata series: Version 5.0 [machine-readable database]. Minneapolis: University of Minnesota.

Smith, J. and F. Welch (1989). Black economic progress after Myrdal. *Journal of Economic Literature*, 519–564.

Vowles, T. (2001). The "Southwest Effect" in multi-airport regions. *Journal of Air Transport Management 7*(4), 251–258.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data.* The MIT press.