ABILITY, SORTING AND WAGE INEQUALITY

Pedro Carneiro
Sokbae Lee

Ability, Sorting and Wage Inequality

Pedro Carneiro and Sokbae Lee*
University College London and Institute for Fiscal Studies

July 2005

Abstract

In this paper we examine the importance of heterogeneity and self-selection into schooling for the study of inequality. Changes in inequality over time are a combination of price changes, selection bias and composition effects. To distinguish them, we estimate a semiparametric selection model for a sample of white males surveyed (during the 1990s) by the National Longitudinal Survey of Youth, but our results are applicable to broader analyses of inequality. In our data, as college enrollment increases in the economy, average college wages decrease and average high school wages increase, and therefore inequality between college and high school groups decreases. Moreover, selection bias causes us to understate the growth of different measures of the average return to schooling in our sample. It also leads us to understate the increase in wage dispersion at the top of the college wage distribution, and to overstate it at the bottom of the college wage distribution.

Keywords: Comparative advantage, composition effects, local instrumental variables, selection bias, semiparametric estimation, wage distribution.

1   Introduction

The rise in wage inequality over the last 30 years in the US has been extensively documented in the economics literature (e.g., Katz and Autor, 1999). Inequality increased both within and between education groups, and these changes have been associated with increases in the return to schooling and ability. Several arguments have been put forward as explanations for these trends such as skill biased technical change, changes in the supply of college graduates or changes in institutions (e.g., Autor, Katz and Kearney, 2004). However, it is not straightforward to infer price changes from observed trends in inequality because individuals are heterogeneous and self-select into different levels of schooling. This implies that measures of within and between group inequality are taken from self-selected samples of individuals, and therefore contaminated by selection bias. Furthermore, increases in wage dispersion were accompanied by increases in the educational attainment of the population. These lead to changes in the composition of different education groups, affecting the trend in wage dispersion. Therefore, bias occurs when trying to assess the determinants of inequality. Observed changes in inequality are a combination of changes in prices, changes in composition and changes in selection bias. In this paper we provide an empirical analysis of the role of heterogeneity and self-selection for the study of inequality.

The literature on the returns to schooling emphasizes the importance of accounting for heterogeneity for both the estimation and interpretation of this parameter (e.g., Card, 1999, Carneiro, Heckman and Vytlacil, 2003). In contrast, most analyses of inequality ignore these concerns and consider instead representative agent models of the labor market (e.g., Katz and Murphy, 1992, Card and Lemieux, 2001). The main reason for neglecting selection is lack of adequate data. Traditionally the analysis of inequality is done with datasets that provide reliable wage information on a large number of individuals over a long period of time (essential for a complete description of the evolution of inequality) but only scarce information on individual characteristics that determine schooling decisions (essential for a complete study of selection).[1] The latter is usually only available in smaller datasets that follow individuals over relatively short periods of time. However, for several economic problems of interest, results from small scale empirical studies are increasingly being used to inform and guide the empirical analysis of more macro trends (or macro models).[2] In this paper we analyze the consequences of heterogeneity and self-selection into schooling for the empirical

---

[1] For example, in the US these are the Census and the Current Population Survey (CPS). In the UK these are the Family Expenditure Survey or the General Household Survey.

[2] See, for example, the discussion in Browning, Hansen and Heckman (1999). While several macro problems have traditionally been analyzed in representative agent frameworks, evidence from micro data shows that heterogeneity is prevalent in a variety of economic problems. In current research heterogeneity has become a central feature of many analyses of macro problems.

study of wage inequality using data from the National Longitudinal Survey of Youth (NLSY) over the 1990s. The NLSY is a nationally representative sample for the U.S. with very rich detail on individual characteristics related to schooling choices and wages, but with a smaller sample size and covering a shorter period than what is used in standard work in this field. Moreover, for all the years we analyze, the wage distribution in the NLSY is similar to the wage distribution in the CPS.[3] From our study we are able to draw general lessons for analyses of inequality.

Using the NLSY we estimate a semiparametric selection model of schooling and wages. We use measures of the costs of school attendance as instruments for schooling: tuition in local four year public colleges, distance to the nearest college and the local unemployment rate in the area of residence in the late adolescent years. Our assumption is that these variables provide exogenous variation in schooling conditional on the extensive set of controls we include in the wage regressions (cognitive ability, parental education, number of siblings and local labor market conditions). The model is estimated using the method of local instrumental variables (Heckman and Vytlacil, 1999, 2001, 2005; Carneiro, Heckman and Vytlacil, 2003). This method was developed for average parameters and we extend it to the estimation of marginal distributions of outcomes in each sector.

In all the years we analyze we find that individuals sort into the level of education where they are comparatively and absolutely better, consistent with Willis and Rosen (1979) and Carneiro, Heckman and Vytlacil (2003). College graduates would have been relatively "low quality" high school graduates had they not gone to college, and high school graduates would have been relatively "low quality" college graduates had they gone to college. As college enrollment increases in the economy we expect average college wages to decrease and average high school wages to increase,[4] and therefore inequality between college and high school groups should decrease.[5] Moreover, selection bias causes us to understate the level and growth of the average return to schooling[6] in our sample. It also leads us to understate the increase in wage dispersion at the top of the college wage distribution, and to overstate it at the bottom of the college wage distribution.

Our findings are of general interest for analyses of inequality. For example (using data from the US Census) the return to college for white males[7] in the 1980s increased from 22% to 33% and the college participation rate went from 41% to 55%. Simulations from our model suggest that in the absence of any increase in college participation the return to college would have increased by

---

[3]For white males born between 1957 and 1964, the demographic group studied in this paper.

[4]Assuming that skill prices are fixed, as in a partial equilibrium framework.

[5]This is just the OLS estimate of the return to college (not including any controls in the regression).

[6]There are different definitions of the average return to college we could consider. Here we refer to the average return for individuals who attend college (usually termed treatment on the treated). In our empirical work we also look at trends in other measures of returns.

[7]Estimated by Ordinary Least Squares (OLS).

7% more during that decade (from 22% to 40%). Furthermore, if the patterns of sorting and the evolution of skill prices in the 1980s were similar to those of the 1990, this estimate is affected by selection bias and understates the growth in the average return to college in the economy.

There exist a few papers that account for heterogeneity and self-selection in the study of inequality, although many of them do not focus on schooling. Most similar to us are Heckman and Sedlacek (1985) and Gould (2001), but they consider selection into industries and occupations. Taber (2001), Chay and Lee (2001) and Chen and Khan (2005) consider selection into schooling with different models than ours. The first two focus on average parameters, while the latter focuses on a cross section and does not systematically consider the role of sorting, selection bias and composition changes. Juhn, Kim and Vella (2005) examine the role of composition changes on the return to schooling using the Census and argue that these can only explain a small fraction of the trend in the OLS estimate of this parameter.[8] Finally, the impact of selection into employment on observed inequality trends has been considered in several papers. Two examples (among many others) are Heckman and Todd (2000) who study the trend in black white wage convergence[9], and Blundell, Reed and Stoker (2003) who focus on aggregate wage growth in the UK.[10] In our paper we ignore selection into employment, which is not likely to be empirically important in our sample.

The remainder of this paper is organized as follows. In section 2 we present an econometric model which underlies our empirical work. In section 3 we apply our model to the study of wage inequality using white males in the NLSY. Using our estimates we document the patterns of sorting of individuals to different levels of schooling and the empirical importance of selection bias and composition effects. Section 4 gives some concluding remarks. In the Appendix we provide a detailed description of the data, and of estimation and testing procedures.

2   Model

Our point of departure is the binary treatment model that is standard in the programme evaluation literature.[11] Let Y_1 and Y_0 be potential individual outcomes in two states, 1 and 0. In this paper

---

[8] Also related is Gosling, Machin and Meghir (2000), who show that most of the evolution of inequality in the UK is explained by cohort effects, and suggest that (among other things) these may be related to changes in composition of different schooling groups.

[9] They show that selective dropout of blacks from employment accounts for most of the recent black-white wage convergence. Chandra (2003) reiterates their conclusion and proposes alternative methods for accounting for the unemployment of blacks. Neal (2004) also shows that accounting for self-selection into unemployment among females changes our characterization of the black-white wage gap among females.

[10] They show that increasing unemployment in the UK accounts for most of the increase in aggregate wage growth since the 1970s. Blundell, Gosling, Ichimura and Meghir (2004) illustrate how different allocations of the unemployed into different sections of the observed distribution of income affects the study of inequality.

[11] The econometric model used in this paper is based on Heckman and Vytlacil (2005) and Carneiro, Heckman and Vytlacil (2003). We consider only two schooling levels, which is a limitation of our analysis, but nevertheless common practice in the literature on inequality. Heckman and Vytlacil (2005) consider models with multiple levels

Y_1 and Y_0 are potential log wages in college and high school for several years in the 1990s.[12] We assume

$$Y_1 = \mu_1\left(X, U_1\right) \quad \text{and} \quad Y_0 = \mu_0\left(X, U_0\right), \tag{1}$$

where X is a vector of observed random variables influencing potential outcomes, μ_1 and μ_0 are unknown functions, and U_1 and U_0 are unobserved random variables. The return to schooling is $\beta = Y_1 - Y_0$ and it varies across individuals.

We assume that individuals choose to be in state 1 or 0 (prior to the realizations of the outcomes of interest) according to the following equation:

$$S = 1 \text{ if } \mu_S\left(Z\right) - U_S > 0, \tag{2}$$

where Z is a vector of observed random variables influencing the decision equation, μ_S is an unknown function of Z, and U_S is an unobserved random variable. In our paper $S = 1$ is college attendance and $S = 0$ is college non-attendance. Equation (2) can be interpreted as the reduced form of a well-specified economic model of college attendance.[13] For each individual, the observed outcome Y is

$$Y = SY_1 + (1 - S)Y_0.$$

The set of variables in X can be a subset of Z. For identification, assume that there is at least one variable in Z that is not in X (exclusion restriction). Furthermore, Z is independent of (U_1, U_0, U_S) given X. As in Heckman and Vytlacil (1999, 2001, 2005), we can rewrite (2) as:

$$S = 1 \text{ if } P > V,$$

where $V = F_{U_S|X}[U_S|X]$, $P = F_{U_S|X}[\mu_S(Z)|X]$, and $F_{U_S|X}(u_s|x)$ is the CDF of U_S conditional on $X = x$. By construction, $V \sim \text{Unif}[0,1]$ conditional on X. Notice that the higher V is, the less likely an individual is to attend a college.

Throughout the paper, for any random vector X, $f_X\left(x\right)$ denotes the PDF of X and $F_X\left(x\right)$ denotes the CDF of X. In addition, for any random variables X and Y, $f_{Y,X}(y,x)$, $f_{Y|X}(y|x)$, and $F_{Y|X}(y|x)$ denote the joint PDF of Y and X and the conditional PDF and CDF of Y on $X = x$, respectively. We suppress subscripts in the notation whenever this can be done without causing confusion.

---

of schooling.

[12] In our empirical work, we follow the NLSY respondents over time and estimate the econometric model described in this section for each year. To simplify presentation, we suppress dependence on time scripts.

[13] Carneiro, Heckman and Vytlacil (2003) use this model to examine heterogeneity in the returns to college and present an economic model that can justify the specification in (2).

2.1 Consequences of Heterogeneity and Selection for the Study of Wage Inequality

Standard analyses of inequality usually start from the following framework:

$$Y = X'\gamma + \varepsilon, \tag{3}$$

where Y is log wage and X is a vector of observable variables, such as the level of schooling and work experience, gender or race. In this model "between group inequality" corresponds to the coefficient(s) on X in the equation (3) (γ), and "within group inequality" refers to the dispersion in ε. A common exercise consists in trying to understand the role of different observable (X) and unobservable (ε) variables in the evolution of wage inequality, as in (for example) Juhn, Murphy and Pierce (1993), DiNardo, Fortin and Lemieux (1996), Autor, Katz and Krueger (1998) or Autor, Katz and Kearney (2004). In this paper we consider a more general setting given by equation (1) where schooling (one of the variables in X) is allowed to be endogenous and there can be different levels of dispersion in different schooling groups (U_1 and U_0).[14] Our goal is to understand the determinants of inequality.

For illustration, suppose we take a simple version of the model of equation (1) where (ignoring X):

$$\begin{aligned} Y_1 &= \mu_1 + \gamma_1 U_1 \\ Y_0 &= \mu_0 + \gamma_0 U_0 \end{aligned} \tag{4}$$

In this model U_1 and U_0 are abilities in each sector and γ_1 and γ_0 are their prices, and we assume that $E[U_1] = E[U_0] = 0$[15] (μ_1 and μ_0 are the two other fundamental prices in this framework). The standard measure of between group inequality (which is usually called the college wage premium) can be written as:

$$\begin{aligned} \beta_{OLS} &= E\left(Y_1|S=1\right) - E\left(Y_0|S=0\right) \\ &= \mu_1 - \mu_0 \text{ (average return to schooling)} \\ &\quad + \gamma_1 E\left(U_1|S=1\right) - \gamma_0 E\left(U_0|S=1\right) \text{ ("sorting gain")} \\ &\quad + \gamma_0 \left[E\left(U_0|S=1\right) - E\left(U_0|S=0\right)\right] \text{ (selection bias).} \end{aligned}$$

---

[14]We also consider a richer set of observable individual characteristics including a measure of cognitive skill and family background, which are usually absent in these studies.

[15]Model (1) is more general than the factor model in (4). It allows X variables to affect μ_1 and μ_0 in a flexible way and it allows for nonlinear relationships between (Y_1, Y_0) and (U_1, U_0).

Several other alternative decompositions are possible.[16] As emphasized in the returns to schooling literature, β_{OLS} does not measure the return to schooling for anyone in the economy (since it compares individuals who are not comparable), and can change over time for reasons which are unrelated with changes in the fundamental prices in the economy (μ_1, μ_0, γ_1 and γ_0), such as changes in the composition of different skill groups (which translate into changes in the distribution of U_1 and U_0 for each group) and changes in the selection bias. Composition effects and selection bias can also affect the evolution of observed within group inequality.

In summary, in order to understand movements in the supply and demand for different skills in the economy one needs to estimate changes in γ_1, γ_0, μ_1 and μ_0, which requires accounting for selection in the analysis of changes in within and between group inequality.[17] Considering these effects is one of the goals of our paper, although we do it in the more general model of equations (1) and (2), where wages can be nonlinear functions of underlying skills. We estimate the evolution of "pricing functions" for observed and unobserved skills over the 1990s, and simulate how changes in educational attainment affect the evolution of inequality.

2.2 Accounting for Selection

It is well known that (in general) the joint distribution of Y_1 and Y_0 cannot be nonparametrically identified in a cross section.[18] Nevertheless, as shown in Heckman and Smith (1998), it is possible to identify $f_{Y_1|X,V}(y_1|x,v)$ and $f_{Y_0|X,V}(y_0|x,v)$, and we estimate these objects in our paper. This section provides our main identification result and section 2.3 gives an informal description of our estimators. Our econometric model builds on Heckman and Vytlacil (1999, 2001, 2005), who develop the Marginal Treatment Effect (MTE from now on, a parameter introduced in the literature by Bjorklund and Moffit, 1987) and use this parameter to unify the literature on treatment effects.[19]

---

[16]For example:

$$
\begin{aligned}
\beta_{OLS} &= E(Y_1|S=1) - E(Y_0|S=0) \\
&= \mu_1 - \mu_0 \text{ (average return to schooling)} \\
&\quad + \gamma_1 E(U_1|S=0) - \gamma_0 E(U_0|S=0) \text{ ("sorting loss")} \\
&\quad + \gamma_1 [E(U_1|S=1) - E(U_1|S=0)] \text{ (selection bias).}
\end{aligned}
$$

[17]Even in a representative agent framework where $\gamma_1 = \gamma_0$ and $U_1 = U_0$, if schooling is endogenous and educational attainments change over time, selection bias and composition changes could still mask the true evolution of prices in the economy. Thus, in general it is not possible to understand the determinants of the evolution of inequality without accounting for selection.

[18]Heckman, Smith and Clemens (1996) compute bounds for $f_{Y_1,Y_0}(y_1,y_0)$ using data from a training program. Carneiro, Hansen and Heckman (2003) estimate $f_{Y_1,Y_0}(y_1,y_0)$ assuming that Y_1 and Y_0 have a factor structure that governs all the dependence between the two variables. They fit models with two independent factors which have a flexible mixture of normals distribution. They develop identification results for models with an arbitrary number of factors. Cunha, Heckman and Navarro (2005) extend this work and estimate models with more than two factors.

[19]Carneiro, Heckman and Vytlacil (2003) apply this framework to an analysis of heterogeneity in the returns to schooling.

We extend their method of local instrumental variables to the estimation of marginal distributions of potential outcomes. Estimation of marginal distributions of potential outcomes has been previously considered by Imbens and Rubin (1997), Abadie (2002) and Abadie, Angrist and Imbens (2002) under the local average treatment effect (LATE) framework of Imbens and Angrist (1994). Our empirical framework draws on Carneiro, Heckman and Vytlacil (2003).

Using the model of Heckman and Vytlacil (1999, 2001, 2005) (as shown in section 2.3), we can construct multiple counterfactuals of interest as weighted averages of these two marginal densities and their functionals. As by-products, we also estimate a variety of treatment parameters. Estimation of counterfactual distributions and treatment effect parameters allows us to gain important insights into the sources of heterogeneity in the labor market, how they determine the main patterns of sorting into schooling and the evolution of the wage distribution.

The following lemma, which is an extension of Heckman and Vytlacil (1999, 2001, 2005), provides identification of our objects of interest in the nonparametric model given by (1) and (2).

Lemma 1 *Consider the nonparametric selection model given by (1) and (2). Let $V = F_{U_S|X}[U_S|X]$ and $P = F_{U_S|X}[\mu_S(Z)|X]$. Assume that (1) $\mu_S(Z)$ is a nondegenerate random variable conditional on X; (2) (U_1, U_S) and (U_0, U_S) are independent of Z conditional on X; (3) The distributions of U_S and $\mu_S(Z)$ conditional on X are absolutely continuous with respect to Lebesgue measure; (4) For a measurable function G, $E|G(Y_1)| < \infty$, and $E|G(Y_0)| < \infty$; and (5) $0 < Pr(S = 1|X) < 1$. Then*

$$E[G(Y_1)|X = x, V = p] = E[G(Y)|X = x, P = p, S = 1] + p\frac{\partial E[G(Y)|X = x, P = p, S = 1]}{\partial p}$$

$$E[G(Y_0)|X = x, V = p] = E[G(Y)|X = x, P = p, S = 0] - (1 - p)\frac{\partial E[G(Y)|X = x, P = p, S = 0]}{\partial p}$$

provided that $E[G(Y)|X = x, P = p, S = 1]$ and $E[G(Y)|X = x, P = p, S = 0]$ are continuously differentiable with respect to p for almost every x.

Proof. Assumptions (1) and (3) ensure that P is a nondegenerate, continuously distributed random variable conditional on X. Assumption (5) guarantees that P is strictly between 0 and 1 conditional on X. Assumption (4) is needed to ensure that expectations considered below are finite. Notice that

$$E[G(Y)|X = x, P = p, S = 1] = E[G(Y)|X = x, P = p, V < p]$$
$$= \int_0^p E[G(Y_1)|X = x, V = v]\, dv/p,$$

where the second equality follows from assumption (2) and the fact that V is uniformly distributed on $[0, 1]$ conditional on X. The first conclusion follows by multiplying both sides of the equation

above by p and differentiating both sides with respect to p. The proof of the second conclusion is similar. ∎

The assumptions in Lemma 1 are basically identical to those of Heckman and Vytlacil (1999, 2001, 2005). The conditional means of Y_1 and Y_0 given $X = x$ and $V = v$ are identified by taking $G(Y) = Y$ and therefore the MTE, defined as $E(Y_1 - Y_0 | X = x, V = v)$, is identified. Furthermore, the conditional distributions of Y_1 and Y_0 given $X = x$ and $V = v$ are identified by choosing $G(Y) = 1(Y \leq y)$, where $1(\cdot)$ is the standard indicator function, and therefore the conditional densities and quantiles are also identified.

Notice that we can only identify $E(Y_1 | X = x, V = p)$ over the support of P for individuals in $S = 1$ conditional on $X = x$, and $E(Y_0 | X = x, V = p)$ over the support of P for individuals in $S = 0$ conditional on $X = x$. As a consequence, we can only identify $E(Y_1 - Y_0 | X = x, V = p)$ over the common support of P for individuals in $S = 1$ and $S = 0$ conditional on $X = x$.

Our procedure has a standard instrumental variables intuition (see also Carneiro, Heckman and Vytlacil, 2003). When we use instrumental variables in a random coefficient model we interpret the estimates as a local average treatment effect or a weighted average of local average treatment effects (see Imbens and Angrist, 1994; Angrist, Graddy and Imbens, 2000; Heckman and Vytlacil, 2005; Carneiro, Heckman and Vytlacil, 2003). Suppose that we have a continuous valid instrument Z (e.g., tuition), and we pick a sample of individuals who have values of the instrument which are very similar, say z_1 and z_2. If we use this subsample to estimate the return to college using standard IV, we identify the average return to college for individuals induced to switch from high school to college when the instrument switches from z_1 to z_2. Similarly, by comparing $E(Y_1 | S = 1, Z = z_1)$ and $E(Y_1 | S = 1, Z = z_2)$ (and $\Pr(S = 1 | Z = z_1)$ and $\Pr(S = 1 | Z = z_2)$) we can identify the average college wages for individuals who change their schooling decision when Z moves from z_1 to z_2 (similarly for high school wages). Finally, from the knowledge of $f(Y_1 | S = 1, Z = z_1)$ and $f(Y_1 | S = 1, Z = z_2)$ it is possible to identify the marginal distributions of college wages for this same group of individuals (and analogously for high school). The latter is the same reasoning behind the results in Imbens and Rubin (1996). These individuals define a particular margin: they are indifferent between going to college or not when the instrument is between z_1 and z_2 (and therefore they switch when the instrument varies over this range). By comparing pairs of z values over the whole observed range of Z we can identify all the objects above for individuals at different margins. This is what we show in Lemma 1.

2.3 Estimating Counterfactual Distributions

The identification result in Lemma 1 is very general since it does not impose any restrictions on the functional forms of μ_1 and μ_0 in (1). However, such a flexible framework has some disadvantages that limit its practical usefulness. One important disadvantage is that the precision of a nonparametric estimator based on Lemma 1 decreases rapidly as the number of continuously distributed components of X increases (curse of dimensionality). Another disadvantage is that it is difficult to have full support of P for some observed values of X, thereby implying that treatment parameters such as ATE, TT, and TUT or counterfactuals distributions such as $f_{Y_1}(y_1)$ and $f_{Y_0}(y_0)$ are not identified. To circumvent these disadvantages in our implementation of Lemma 1, we specify and estimate a separable version of (1)[20] under a more stringent assumption on unobservables, but one that is relatively standard in empirical work: we assume that (U_1, U_0, U_S) are independent of (Z, X).[21] In addition, we impose flexible but parametric forms for μ_1, μ_0, and μ_S on the model so that estimating the model reduces to a feasible semiparametric estimation problem.[22]

We first consider estimation of $E[Y_1|X = x, V = v]$ and $E[Y_0|X = x, V = v]$. Under the assumption that (U_1, U_0, U_S) are independent of (Z, X),

$$E[Y_1|X = x, V = v] = \mu_1(x, \beta_1) + E[U_1|V = v],$$

and

$$E[Y_0|X = x, V = v] = \mu_0(x, \beta_0) + E[U_0|V = v],$$

where the functional forms of μ_1 and μ_0 are specified up to finite dimensional parameters β_1 and β_0. Thus, estimates of $E[Y_1|X = x, V = v]$ and $E[Y_0|X = x, V = v]$ can be obtained by estimating β_1, β_0, $E[U_1|V = v]$, and $E[U_0|V = v]$.

We estimate β_1 and β_0 using a semiparametric version of the sample selection estimator of Das, Newey, and Vella (2003). Notice that under the assumption that U_1 and V are independent of X and Z, we have

$$E[Y|X = x, P = p, S = 1] = \mu_1(x, \beta_1) + \lambda_1(p), \tag{5}$$

where $\lambda_1(\cdot)$ is an unknown function of P. Equation (5) suggests that β_1 can be estimated by

---

[20]The assumption of separability implies the following modification in our model:

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0,$$

as opposed to $Y_1 = \mu_1(X, U_1)$ and $Y_0 = \mu_0(X, U_0)$.

[21]In contrast, Lemma 1 only requires that (U_1, U_0, U_S) are independent of Z given X, which allows for dependence between (U_1, U_0, U_S) and X. The separability and independence assumptions we impose are standard in most empirical work in economics.

[22]$\mu_1(X)$, $\mu_0(X)$ and $\mu_S(Z)$ will be functions of second degree polynomials and pairwise interactions between their arguments. Exact specifications are given in section 3.1.

a partially linear regression of Y on X and P using only observations with $S = 1$. Since P is unobserved, Das, Newey, and Vella (2003) suggest a two-step procedure. The first step consists in the construction of the estimated P and the second step consists in the estimation of β_1 using the estimated P. In this paper, the first step is carried out by a logit regression of S on Z using a flexible but parametric form of $\mu_S(z)$, and the second step is accomplished using a Robinson (1988)-type estimator with the estimated P.[23] Analogously, β_0 can be estimated by a partially linear regression of Y on X and estimated P using only observations with $S = 0$.

Once we estimate β_j and $\lambda_j(p)$, estimation of $E[U_j|V = v]$ for $j = 0, 1$ comes from directly applying Lemma 1 with $G(u) = u$:

$$E(U_1|V = v) = E[U_1|P = v, S = 1] + v\frac{\partial E[U_1|P = v, S = 1]}{\partial p} \tag{6}$$

$$E(U_0|V = v) = E[U_0|P = v, S = 0] - (1 - v)\frac{\partial E[U_0|P = v, S = 0]}{\partial p}. \tag{7}$$

Equations (6) and (7) are the basis for nonparametric estimators of $E[U_1|V = v]$ and $E[U_0|V = v]$ proposed in this paper. Local polynomial estimation is used here to estimate $E(U_1|P = v, S = 1)$ (which corresponds to $\lambda_1(p)|_{p=v}$ in equation (5)), $E(U_0|P = v, S = 0)$ and their partial derivatives with respect to P. This is because local polynomial estimation not only provides a unified framework for estimating both a function and its derivative but also has a variety of desirable properties in comparison to other available nonparametric methods. Fan and Gijbels (1996) provide a detailed discussion of the properties of local polynomial estimators. In Appendix B.1 we present a detailed description of the method for estimating $E[U_1|V = v]$ and $E[U_0|V = v]$.

Finally, notice that $f_{Y_1|X,V}(y_1|x, v)$ and $f_{Y_0|X,V}(y_0|x, v)$ can be obtained by location shifts from $f_{U_1|V}(u_1|v)$ and $f_{U_0|V}(u_0|v)$, i.e.,

$$f_{Y_1|X,V}(y_1|x, v) = f_{U_1|V}(y_1 - \mu_1(x, \beta_1)|v) \quad \text{and}$$

$$f_{Y_0|X,V}(y_0|x, v) = f_{U_0|V}(y_0 - \mu_0(x, \beta_0)|v).$$

To obtain $f_{Y_1|X,V}(y_1|x, v)$ and $f_{Y_0|X,V}(y_0|x, v)$, once we know β_1 and β_0 we only need to estimate $f_{U_1|V}(u_1|v)$ and $f_{U_0|V}(u_0|v)$. As in (6) and (7), we can obtain identifying relationships for $f_{U_1|V}(u_1|v)$ and $f_{U_0|V}(u_0|v)$ and resulting sample analog estimators can be constructed. Note that given estimators of PDF's, it is straightforward to obtain estimators of corresponding CDF's by integrating the estimated PDF's, and to obtain estimators of corresponding quantiles by inverting the estimated CDF's (see Appendix B.2 for details).

---

[23]Series estimation is used in Das, Newey, and Vella (2003) for both the first and second steps. See also Heckman, Ichimura, Smith and Todd (1998).

Heckman and Vytlacil (1999, 2001, 2005) show how we can construct a variety of treatment effect parameters as weighted averages of $E(Y_1 - Y_0 | X = x, V = v)$, and develop weights for several parameters of interest. Drawing on their work, we can estimate $E[Y_j]$, $E[Y_j | S = 1]$, and $E[Y_j | S = 0]$ by integrating out our estimator of $E[Y_j | X = x, V = v]$ with some suitable weights for $j = 0, 1$. Specifically, we obtain estimators of $E[Y_j]$, $E[Y_j | S = 1]$, and $E[Y_j | S = 0]$ by the sample analogs of the following formulae:

$$E[Y_j] = \int \int_0^1 E[Y_j | X = x, V = v] f_X(x) \, dv \, dx,$$

$$E[Y_j | S = 1] = \int \int_0^1 E[Y_j | X = x, V = v] \frac{1 - F_{P|X}(v|x)}{\Pr(S = 1)} f_X(x) \, dv \, dx, \tag{8}$$

and

$$E[Y_j | S = 0] = \int \int_0^1 E[Y_j | X = x, V = v] \frac{F_{P|X}(v|x)}{\Pr(S = 0)} f_X(x) \, dv \, dx.$$

for $j = 0, 1$. See Appendix B.3 for a detailed description of the estimation procedure. Using estimates of these conditional expectations, standard treatment effect parameters can be estimated: ATE $= E[Y_1] - E[Y_0]$, TT $= E[Y_1 | S = 1] - E[Y_0 | S = 1]$, TUT $= E[Y_1 | S = 0] - E[Y_0 | S = 0]$, and OLS $= E[Y_1 | S = 1] - E[Y_0 | S = 0]$. $E[Y_1 | S = 1]$ and $E[Y_0 | S = 0]$ can also be estimated directly by taking sample means of observed college and high school wages. Therefore, comparison between model-based and direct estimates of $E[Y_1 | S = 1]$ and $E[Y_0 | S = 0]$ provides a goodness-of-fit check of our model. Similarly, integrating our estimators of $f_{Y_j | X, V}(y_j | x, v)$ for $j = 0, 1$ with the weights in (8), we can obtain estimators of $f_{Y_j}(\cdot)$, $f_{Y_j | S=1}(\cdot | S = 1)$, and $f_{Y_j | S=0}(\cdot | S = 0)$ for $j = 0, 1$. We use counterfactual wage distributions of Y_1 and Y_0, namely $f_{Y_1}(\cdot)$ and $f_{Y_0}(\cdot)$, to study the evolution of inequality in the economy accounting for self-selection into schooling. Note that $f_{Y_1 | S=1}(\cdot | S = 1)$ and $f_{Y_0 | S=0}(\cdot | S = 0)$ can also be estimated directly by taking sample analogs of observed college and high school wages, which again allows us to do a goodness-of-fit check of our model.

3   Selection Bias, Composition Effects and the Evolution of Inequality

3.1   Data and Specification of the Model

The dataset we use consists of a sample of white males surveyed in the NLSY. Most analyses of the evolution of inequality in the US are based on the CPS (for a recent example see Autor, Katz and Kearney, 2004). The CPS is an annual, representative sample of the entire US population with a large sample size. However, the relative lack of detail on individuals in the CPS does not allow us to properly model the process of selection into schooling and its influence on wages. For this purpose,

the NLSY is considerably better because the NLSY provides much richer individual data than the CPS.[24] In the NLSY there exists detailed information on cognitive ability and family background, which are important determinants of both schooling and labor market outcomes. Furthermore we know the place of residence of most respondents in the NLSY during their adolescent years. As a result, we can construct school and labor market characteristics in different areas of residence of adolescent NLSY respondents and use them as instrumental variables for schooling.[25] The reason we choose to start our analysis in the 1990s and not before is because NLSY respondents were very young in the 1980s.[26] Unfortunately we are unable to separate age and time effects in our analysis, but this problem is not specific to our study.[27] We estimate the model for 1992, 1994, 1996, 1998 and 2000.

Our sample consists of white males born between 1957 and 1964. The hourly wage measure we use was created by the NLSY. In order to minimize measurement error and reduce concerns with selective unemployment, our wage measure for each year is a 5 year average of all non-missing wages reported in the five year interval centered in the year of interest.[28] The model of section 2 only allows for selection into two levels of schooling, so we need to group some schooling categories into these two. The two groups we consider are: high school graduates plus high school dropouts; and some college plus college graduates and above. The main reason for doing this aggregation in our paper is the lack of the credible exclusion restrictions needed to estimate a model with multiple levels of schooling (see Heckman and Vytlacil, 2005). Furthermore, this type of aggregation is very common in the literature on wage inequality where often only two schooling groups are considered.[29]

---

[24]Descriptions of the NLSY and the CPS are available from the Bureau of Labor Statistics: http://www.bls.gov. See also appendix A for a description of the data we use in the paper.

[25]However, the NLSY has three disadvantages: it surveys a very limited set of cohorts (individuals born between 1957 and 1964), the sample size is much smaller than in the CPS, and it is only collected bi-annually after 1994. Therefore, it is important to know whether the patterns of inequality in the NLSY for the 1990s are similar to the patterns of inequality in the CPS for the same cohorts and same time periods. In appendix we show that we can roughly replicate the wage distribution in the CPS during the 1990s for the cohorts born between 1957 and 1964 using the NLSY (see figure 11, which shows percentiles 10, 25, 50, 75 and 90 of the CPS and NLSY wage distribution for each year we study). The NLSY has one more important advantage in addition to the richness of the data: it is a longitudinal study. However, we do not emphasize this advantage in this paper because we only use the longitudinal structure of the NLSY to construct our measure of wages.

[26]Even though many of our respondents are still relatively young during the period of our analysis, we wanted to avoid using measures of labor market outcomes very early in each individual's career since this is known to be a very turbulent period.

[27]Furthermore, even if a large part of the price changes in our data are due to age effects, it is still interesting to correct these for selection. For the purposes of our analysis (although not for the purposes of interpretation), it does not matter which forces are behind the changes in skill prices.

[28]The percentage of individuals in our sample who have a missing observation for our measure of wages (due to unemployment or non-reporting) is the following for each year: 4.1% in 1992, 3.8% in 1994, 3.6% in 1996, 2.8% in 1998 and 2.8% in 2000 (on top of this there is attrition in the panel, which we also ignore in the paper). When we use different measures of wages such as yearly wages or averages over three years of wages, our results are qualitatively similar but they are more imprecise.

[29]This type of aggregation is also common in studies of the returns to schooling using selection models, such as Willis and Rosen (1979) or Carneiro, Heckman and Vytlacil (2003). We start by defining four schooling categories: high school dropouts, high school graduates, some college and college graduates. Because there are multiple useful reports of schooling in the NLSY we construct the educational categories as follows: individuals without a high school

12

We now turn to the exact specification of the equations that we estimate. The X vector in the log wage equations includes years of actual experience[30], the Armed Forces Qualifying Test score (AFQT, a measure of cognitive ability)[31], number of siblings, mother's years of schooling, father's years of schooling, and the state unemployment rate in the current state of residence (five year average centered in the year of interest). Each variable enters with a linear and a quadratic term. We also interact number of siblings, mother's education and father's education. Finally, we include a dummy variable for being a high school dropout, another dummy variable for being a college attendee without a college degree, and interactions of these variables with quadratic polynomials in experience and AFQT. This is an attempt to allow for some selection on observables within each broad schooling category.

The Z vector in the school choice equation consists of AFQT, number of siblings, mother's education, father's education, distance to any college at age 14, average tuition in four year public colleges in the county the individual lived in at age 19[32], and the unemployment rate in the state of residence in 1979.[33] The variables that we exclude from the outcome equations are distance to college, tuition, and local unemployment rate. AFQT and family background enter the schooling choice equation but do not play the role of an instrument since they are included in the X vector as well. In the school choice model, all variables enter with a linear and a quadratic term and are interacted with each other. Our flexible but parsimonious specification of the outcome and choice equations does not restrict the patterns of sorting allowed for our model and ensures that our results are driven by genuine variation in the data, and not by the nonlinearity of the selection equation.

The instrumental variables we use for identification of the model (exclusion restrictions) are intended to measure different costs of attending college and are based on the geographic location of individuals in their late adolescence (and they are essentially the same instruments used in Carneiro, Heckman and Vytlacil, 2003). If the decision of going to college and the (prior) location decision are correlated then our instruments may not be valid (if the determinants of location are correlated with wages). For example, individuals who are more likely to enroll in college may choose to locate in

---

degree are high school dropouts; individuals with a high school degree but with less than 13 years of schooling are high school graduates; those reporting 13 to 15 years of schooling and without a four year college degree go into the some college group; finally, those reporting a four year college degree or 16 or more years of schooling are considered to be four year college graduates. GED recipients who never attend college are included in the group of high school graduates.

[30]Measured in weeks of actual work experience (then converted into years of experience by dividing by 52).

[31]We adjust the AFQT score by the amount of schooling each individual has at the time they take the test (see Hansen, Heckman and Mullen, 2004, and appendix A).

[32]Alternatively we could use the county of residence at age 17. The results are essentially the same but the instrument is less strong, so we prefer to present the specification with tuition at 19.

[33]We experimented other instruments such as distance to four year college only (and other measures of distance at different ages), tuition in two year colleges (and other measures of tuition at different ages), and other versions of the unemployment rate, but these are weaker predictors of college attendance than the variables we decided to use.

areas where colleges are abundant and inexpensive. However, in our wage equations we control for measured ability and several family background variables (mother's years of schooling, father's years of schooling and number of siblings). Therefore, our assumption is that our instruments are valid conditional on measured ability and family background variables, which are also correlated with location choice. Distance to college was first used as an instrument for schooling by Card (1993) and was subsequently used by Kling (2001) and Cameron and Taber (2004). Carneiro and Heckman (2002) show that distance to college in the NLSY79 is correlated with a measure of ability (AFQT), but in this paper we include this measure of ability in the outcome equation. Tuition was used by Kane and Rouse (1995) (although not in the published version of their paper). Average tuition in the county of residence may also be a problematic instrument if it is correlated with average college quality in the county. Notice, however, that we only consider 4-year public institutions, and therefore, exclude all private universities for whom the link between price and quality may be stronger. Again, our assumption is that the included ability and family background are enough to control for the individual and family choice of college quality. Finally, local labor market variables have been used by Cameron and Taber (2004). However, they use a measure of local wage, instead of a measure of local unemployment. They also control for long term wages in the county of residence both in the selection and in the outcome equations, so that the instrument measures business cycle fluctuations orthogonal to the long term quality of the location of residence. Finally, they use county level local labor market variables while we use state level variables. In the outcome equations that we estimate we include the state unemployment rate in the year in which wages are measured.[34]

Sample statistics are presented in table 1. In each year of our data, individuals who attend college have on average higher wages than those who do not attend college. They also have higher levels of cognitive ability, fewer siblings, more educated parents, live nearer to colleges, in counties with lower average tuition and in states with lower unemployment rates in year 1979 than those individuals who never enrolled in college. It can be seen that college enrollment rates increase from 50% (1992) to 53% (2000), although we follow relatively mature cohorts for college enrollment (the NLSY respondents in year 1992 are between 28 and 35 years old).[35]

---

[34]Furthermore, our estimates of the coefficients on the observables in the outcome and selection equations are roughly robust to the inclusion of dummy variables for state of residence at 14 in the selection and outcome equations, although they become more imprecise. The largest differences between models with and without state dummies are observed in the outcome equations for 1998, and in the effect of unemployment on the selection equation for 1996-2000, but even these ones are not substantial (these tables are available on request). Including such a dummy removes time invariant effects of the state of residence in adolescence, such as a permanent state college quality effect or higher education policy concern, or permanent state labor market conditions.

[35]If we use this data to produce standard OLS and IV estimates of the returns to college we get estimates roughly in line with the literature (although estimates of returns to college are not very common). Using our data we run the following regression:

$$Y = \alpha + \beta S + \gamma' X + u$$

14

We use a logit model for schooling choice.[36] The flexible logit model used here provides good estimation precision, allows for nonlinear effects of Z, and ensures that the estimated probability lies between 0 and 1. Furthermore, since the dependence between V and (U_1, U_0) is not constrained in any way, assuming this parametric form for the choice equation is probably not restrictive. We estimate one logit model for each year.[37] Average derivatives are presented in table 2. Ability and family background are strong predictors of college attendance. Tuition is also an important determinant of enrollment in college. Even though distance and local unemployment do not strongly predict college attendance on average, we choose to include them in the model because they potentially play a useful role in the expansion of the support of P and may affect college enrollment of some specific groups of individuals, although not for all individuals.

The identification of the objects of interest was discussed in section 2.2. The exact procedure which we use to estimate this model is described in detail in appendix B. For each year we estimate $f_{Y_1|X,V}(y_1|x,v)$ and $f_{Y_0|X,V}(y_0|x,v)$ and then weight these objects with appropriate weights to construct the counterfactuals of interest, as described in section 2.3. However, it is only possible to estimate these functions within the support of the data. In particular, we can only estimate them for values of X and P (accordingly V) for which we have individuals both in the college and high school groups. Figure 1 shows the support of the data for 1994, a representative year in our sample.[38] The top two figures refer to P and the bottom two figures refer to AFQT. AFQT is only one of the variables in the X vector on which we condition, but it is the most important one and is also most likely to have non-overlapping supports among X (see table 1).[39] Notice that the support of P is almost the full unit interval which allows us to estimate our model over the full support of V. We are able to achieve large support for P because: (i) we combine multiple instruments into an index; more importantly, (ii) if we assume that X is independent of (U_1, U_0, V) we can trade-off variation in X and Z to increase the support of P (since X is controlled for in the outcome equations in a

where Y is our measure of log hourly wages, S is a dummy for college attendance and X includes all the Xs we include in our own outcome equations, entering with linear, quadratic and interaction terms. OLS estimates of β go from 0.236 in 1992 (standard error of 0.05) to 0.372 in 2000 (standard error of 0.05), while IV estimates go from 0.902 (0.344) to 0.997 (0.336) over the same period. For comparison, using least squares and much less controls than us, and using weekly wages (from the CPS) instead of hourly wages (for 1993-1997, a comparable time period), Welch (1999) estimates (by OLS) a dropout-high school premium of 30-33%, a some college-high school premium of 24-35% and a college graduate-high school premium of 84-96%.

[36] The propensity score $P(Z)$ could be estimated nonparametrically; however, dimension reduction is needed here to achieve reasonable precision of estimates since the dimension of Z is large.

[37] Alternatively we could have estimated a single selection model for all the years of the sample. The reason we choose not to do it is that, even though these individuals are well into their adult years in the beginning of the 1990s, there are still changes in schooling attainment during the decade. In particular, the college enrollment rate in this sample increases from 50% to 53%. A similar pattern is found in the CPS. When we redo the analysis considering that schooling is fixed at a particular level for all the years the overall results do not change substantially.

[38] Other years are available from the authors on request. This figure changes very little across years.

[39] On results available on request, we show that if we show similar figures for an index of X variables (namely, $E[\mu_S(X,Z)|X]$) we get roughly the same conclusions.

15

very flexible way).[40]

3.2 Empirical Results

There are three components in our empirical analysis. First, we characterize how individuals sort into different levels of schooling and illustrate how sorting affects inequality. Second, we study the determinants of inequality and of its evolution over time. Third, we investigate the role of composition changes in the evolution of inequality. We also characterize the effect of education policy on wage inequality in a partial equilibrium framework.[41]

3.2.1 Characterizing the Patterns of Sorting

We start by presenting our estimates of $E(Y_1|X,V)$ and $E(Y_0|X,V)$ for 1994, a representative year in our sample. Figure 2 shows estimates of $E(Y_1|AFQT,X,V=0.5)$, $E(Y_0|AFQT,X,V=0.5)$, and $E(Y_1-Y_0|AFQT,X,V=0.5)$, as functions of AFQT, along with 90% bootstrap confidence intervals for $E(Y_1-Y_0|AFQT,X,V=0.5)$. We fix years of experience at 10 to abstract from life-cycle effects, V at its median value, and the remaining variables in X at: 3 siblings, 12 years of mother's and father's education, and 7% of local unemployment rate. In figure 3 we graph $E(Y_1|AFQT=0,X,V)$, $E(Y_0|AFQT=0,X,V)$, $E(Y_1-Y_0|AFQT=0,X,V)$ (the MTE of Heckman and Vytlacil, 1999, 2001, 2005), as functions of V, along with 90% bootstrap confidence intervals for $E(Y_1-Y_0|AFQT=0,X,V)$. Again we fix years of experience at 10 and the remaining X variables at the values described above apart from AFQT, which we fix at its mean value 0. We focus on AFQT and V because they are both strong determinants of college attendance and they are strongly correlated with wages. The wage-AFQT gradient is steep in college but not in high school, and as a consequence the returns to college increase with AFQT. As for V, it has a negative correlation with college wages and a positive correlation with high school wages. Recall that the higher the V is, the smaller the likelihood that an individual enrolls in college.

For every year in our sample we find strong evidence favoring existence of selection on unob-

---

[40]The intuition for this can be given by the following example: variation in Z for individuals with high levels of X (say AFQT) will mostly affect individuals with high levels of V, while variation in Z for individuals with low levels of X will affect mostly individuals with low levels of V. This assumption is standard in the empirical literature but its use is not made as explicit as here. This procedure has some similarities with the use of interactions between X and Z as instruments for the endogenous variables (controlling for X in the outcome regression). It is also often used in selection models (although this is rarely stated explicitly). In our application, this independence assumption plays an important role in the practical implementation of the method of local instrumental variables since it allows us to extend the support of the instrument. Since we are very flexible in the way we include X in the wage regression we can be confident that this type of expansion in the support of the instrument does not compromise its validity, provided that both X and Z are independent of (U_1,U_0,U_S).

[41]In the appendix we carry out an informal goodness-of-fit of our model by comparing objects that can be estimated using the model and directly from the data. As shown in figure 12 and explained in appendix C, our model fits the data considerably well. This result was not guaranteed to happen given the multiple step procedure we adopt, and therefore it gives us confidence in the specification of the model.

servables for college wages and selection on AFQT, family background, and unobservables for both college and high school wages using simple Hausman-type tests. These results suggest that the patterns reported above and throughout the paper are genuinely important in the data, and therefore we need to account for heterogeneity and selection in analyses of inequality. A detailed description of the tests performed and of the results can be found in appendix C.

Our results confirm the findings in Willis and Rosen (1979), Carneiro, Heckman and Vytlacil (2003), and Carneiro and Lee (2004). Those individuals most likely to enroll in college (the ones with high levels of AFQT and low levels of V) have high wages if they decide to go to college (since college wages increase with AFQT and decrease with V) but have low wages if they decide not to go to college (since high school wages do not move substantially with AFQT and increase with V). Conversely, individuals less likely to enroll in college have low college wages and high high school wages. Individuals sort into the sector where they have both comparative and absolute advantage. Single skill models of the labor market implicit in standard specifications of earnings equations with no heterogeneity predict college individuals to have higher earnings both in the high school and college sectors than high school individuals. Our findings are inconsistent with such a model.

These results have important implications for the study of inequality which we develop in sections 3.2.2 and 3.2.3. First they indicate that selection bias may play an important role in standard estimates of the returns to schooling and of the effects of different variables on wage inequality. Second, they suggest that, if anything, both the observed high school and college wage distributions may be "missing" their lower tails, since the "worst" individuals in each sector choose to enter the other sector. If we observe increases in educational attainment in the population (composition changes), they are likely to be caused by individuals who (on average) leave the bottom of the high school wage distribution and enter (on average) the bottom of the college wage distribution. As a result, changes in the educational attainment of the population can induce changes in within and between group inequality even in the absence of changes in the price of skill.

Similar patterns are found for other years, although with some differences (the absolute advantage pattern is basically always present in one form or another). In figure 4 we plot these curves for 1992 and 2000, the first and last years that we analyze.[42] There is an upward sloping relationship between AFQT and college log wages for all the years in the sample, which moves up and becomes flatter over time. This indicates an overall rise in wages (either over time or over the life-cycle), but a decrease in the slope of the AFQT-Wage gradient. If the slope of this function could be interpreted as the price of AFQT, this movement would be analogous to a fall in the price of AFQT in college. In

[42]In the appendix we present the figures for all the years we analyze.

17

high school, log wages are slightly increasing in 1992 but they have an inverse U-shape relationship with AFQT for 2000.[43] In terms of the V-Wage gradient, heterogeneity seems to be quantitatively important in high school for both 1992 and 2000, but in the college sector once again this function flattens out over time.[44] In the appendix we present average marginal effect of each observable variable on high school and college wages which confirm what we see in the pictures.[45]

Although we tentatively referred to these functions as pricing functions in the discussion above, it is not completely clear how to interpret them. Heckman and Sedlacek (1985) interpret them as production functions of sector specific tasks, and the price of the task is given by the intercept of the function.[46] Alternatively, we could interpret these as pricing functions for individual attributes X and V, and we allow for nonlinear pricing of attributes as well as sector specific prices. However, apart from AFQT, it is not clear whether the variables in X and V are relevant measures of skill.[47] It is more likely that they proxy relevant skills, or are inputs in the production of relevant skills, as in a version of the approach of Heckman and Sedlacek (1985) with multiple skills within each sector. Probably we need a mixture of both models.[48]

There may also be unobserved skills whose prices also change over time, but on which agents do not select when making their schooling decisions, and which we have ignored so far. Even after we condition on X and V (the sources of selection in our model) there is substantial dispersion in Y_1 and Y_0. Carneiro, Hansen and Heckman (2003) and Cunha, Heckman and Navarro (2005) interpret

---

[43]Recall that due to limitations of the data we aggregate high school dropouts and high shool graduates in the same sector, which may give rise to some of the patterns we observe. Say low AFQT individuals are high school dropouts, high AFQT individuals enroll in college and the remaining are high school graduates. Then, if forced into the high school sector, high school dropouts would have low wages as would college graduates, while actual high school graduates would have the highest wages. This may partly explain why such a U-shaped pattern can emerge. Carneiro and Lee (2004) exclude dropouts from the sample and get a strong pattern of absolute advantage both in terms of observables and unobservables.

[44]In 2000 there is an upward sloping section in the expectation of Y_1 as a function of V that we cannot explain, but which can be due to sampling error. However, for most of its range the function is downward sloping, as in 1992.

[45]Notice that even though the average derivatives reported in the appendix do not seem to be very precisely determined, we can test and reject that these variables do not belong in the wage equations, as reported in appendix C. Furthermore, since some of these functions are non-monotonic the average derivatives reported here may not always be interesting objects to examine.

[46]Even though this is a very appealing interpretation, a change in the price of sector specific task cannot lead to an increase in within group log wage inequality, which is not an attractive feature in studies of the evolution of inequality since it is a restriction at odds with the data. Furthermore, since we allow not only the intercept but the whole function to vary across different years, this would imply that the production function can change over time and changes in prices could only be distinguished from changes in technology through arbitrary normalization.

[47]Furthermore, since V represents everything that is unobservable and determines selection, it is likely to be an index of multiple variables.

[48]As emphasized above, the problem is that in that case it is not possible to distinguish changes in technology from changes in prices. For simplicity, if we assume that the technology of skill specific skills is fixed then changes in these functions are given by changes in prices (even though in any given year it is not possible to distinguish prices of skills and the production function of skills). In such a setting, our results could indicate that the prices of the skills associated with characteristics X and V are declining over time. At the same time, we also observe a rise in the overall level of wages (these functions shift up over time), especially in the college sector. So, more specifically, our results indicate an increase in the average price of different skills but a decline in their marginal price. However, even though we attempt to fix experience when plotting these figures, in reality we cannot hope to distinguish experience from time effects. This increase in wages may just be reflecting the effect of experience on wages.

$E(Y_1|X, V)$ and $E(Y_0|X, V)$ as *ex-ante* heterogeneity and $f(y_1|x, v)$ and $f(y_0|x, v)$ as uncertainty faced by the agents at the time they are making the college decision (since agents do not select on it). As emphasized in these papers, the magnitude of uncertainty is very large. In table 3 we compute the variance of counterfactual wages in each sector for all the years and decompose it into a component due to X, a component due to V and the residual component that we emphasize in this paragraph.[49] It is clear that it is the latter that accounts for most of the variance in both high school and college wages, and that its importance increases dramatically over time in both sectors, although it declines slightly in high school for the last two years of the sample. This has important implications both for the interpretation of inequality patterns and for our study of selection bias and composition effects. For example, the R-square of standard wage regressions is usually low indicating that education explains a small amount of the variance of wages and therefore education policy is not likely to strongly affect wage inequality. Table 3 suggests that even if we include in this regression all of the observable and unobservable variables correlated both with schooling and wages (often referred to as ability) this conclusion will not change dramatically.

Figure 5 presents estimates of the 25th, 50th and 75th percentiles of $f(u_1|v)$ and $f(u_0|v)$ for 1992 and 2000 (which correspond up to location to $f(y_1|x, v)$ and $f(y_0|x, v)$ since we do not allow the dispersion of log wages conditional on X to vary with X). In this figure, U_1 and U_0 are normalized to have mean zero. These three quantiles vary with V in the same way average wages in each schooling level vary with V. While there is a non-monotonic relationship between the dispersion in Y_1 and V (which is flat over a large range of V in 1992) the dispersion in Y_0 is clearly increasing with V. This indicates that the components of heterogeneity that do not determine selection are either more disperse for individuals with a higher level of V (indicating more uncertainty in high school wages as an individual becomes more and more likely to belong to the high school sector) or the prices of the skills underlying this heterogeneity are higher for individuals with a high V. Notice also that it is implicit in table 3 that the dispersion in $f(y_1|x, v)$ and $f(y_0|x, v)$ increases from 1992 to 2000. Therefore, even though this type of heterogeneity does not determine selection in the same way X and V do, we observe that individuals with more uncertain wages both in high school and in college select into high school. One interpretation is that there is a large overlap in the high school and

---

[49]In particular:

$$
\begin{aligned}
\mathrm{Var}(Y_1) &= \mathrm{Var}[\mu_1(X)] + \mathrm{Var}(U_1) \\
&= \mathrm{Var}[\mu_1(X)] + \mathrm{Var}[E(U_1|V)] + E[\mathrm{Var}(U_1|V)],
\end{aligned}
$$

where $\mathrm{Var}[E(U_1|V)]$ is the component of variance due to V and $E[\mathrm{Var}(U_1|V)]$ is the remainder. $\mathrm{Var}(Y_0)$ can be decomposed in the same way. The first row of table 3 corresponds to $E[\mathrm{Var}(U_1|V)]$, the part of the variance that cannot be associated with selection, the second corresponds to $\mathrm{Var}[\mu_1(X)]$, the third corresponds to $\mathrm{Var}[E(U_1|V)]$ and the fourth corresponds to $E[\mathrm{Var}(U_1|V)]/\mathrm{Var}(Y_1)$, the fraction of the variance in Y_1 that is not due to any variable related to selection into schooling.

college wage distributions, and for those individuals who face larger dispersion of high school wages it may pay off not to undertake the cost of going to college because they can achieve a high wage in the high school sector, if they are "lucky". Alternatively, we can interpret individuals with high V as high school dropouts (since we group high school graduates and high school dropouts in the same schooling group) and they can be riskier individuals for many possible reasons. Of course, there are other possible interpretations for this pattern.[50]

3.2.2 The Importance of Selection Bias

There exist only small changes in educational attainment over time (composition changes) in our sample since we follow a particular cohort, namely NLSY respondents. Nonetheless, as we show in this section, even in the absence of composition changes the pattern of self-selection documented in section 3.2.1 has important consequences for the study of the evolution of the return to schooling and for the study of the determinants of within group wage inequality.

Figure 6 displays the evolution of $E(Y_j)$ and $E(Y_j|S=j)$ for $j=0,1$.[51] Notice that on one hand, there are large differences in the level as well as the evolution over time between counterfactual and observed means of college wages and on the other hand, there are relatively small differences between counterfactual and observed means of high school wages, although they are not zero. This suggests that selection bias may be quantitatively more important in the college sector than in the high school sector.[52] As a result, the trend in the OLS estimate of the college premium does not correspond to the trend in more meaningful measures of the returns to schooling. In figure 7 we plot the OLS estimate of the college premium ($E(Y_1|S=1) - E(Y_0|S=0)$), and estimates of $E(Y_1 - Y_0)$ (average treatment effect, or ATE), $E(Y_1 - Y_0|S=1)$ (treatment on the treated, or TT) and $E(Y_1 - Y_0|S=0)$ (treatment on the untreated, or TUT) over the 5 years we study. There are two measures of OLS: the one based on our model and the other using data directly. They are quite similar in both the level and trend, implying that we fit the data well. The OLS parameter based on the data directly increases from 35% (in 1992) to 47% (in 2000). However, as emphasized in section 2.1 and in the whole literature on the returns to schooling, OLS does not correspond to the return to schooling for anyone in the economy. As shown in Carneiro, Heckman and Vytlacil (2003) and confirmed in our work for several years of data, the OLS parameter understates the level of TT

---

[50]We need to acknowledge that our model has a limitation since it does not allow the dispersion of log wages to vary with X, something that has been emphasized in recent papers such as Lemieux (2004) or Autor, Katz and Kearney (2004). Nevertheless, the main X variables considered in those papers are education, experience and institutional variables. We ignore the latter and allow the variance of wages to vary with education, so the restriction is only obvious for experience (a variable that does not determine selection).

[51]The evolution of $f_{Y_j}(y_j)$ and $f_{Y_j|S=j}(y_j|S=j)$ for $j=0,1$, is available on request.

[52]This is consistent with results of tests for selection reported in appendix C.

and overstates the level of TUT. The estimation of OLS relies on the comparison of non-comparable individuals. Individuals who do not enroll in college have higher high school wages than average college participants would have had they not enrolled in college ($E[Y_0|S = 0] > E[Y_0|S = 1]$). Therefore, the former overstate the high school wages of the latter. In consequence the estimated return for these individuals is understated (OLS < TT). Conversely, those who do not attend college would be poor college individuals ($E[Y_1|S = 0] < E[Y_1|S = 1]$), and therefore the comparison group we use grossly overstates their opportunities as college graduates, leading to an overstatement of the return (OLS > TT).[53] We observe that TT is declining from 1992 to 1994 but it rises by 35% over the rest of the period, ATE increases by 39% from 1992 to 2000 and TUT increases by 49% over the same period. The trend in the OLS estimate of the return to schooling plays a fundamental role in modern characterizations of the labor market and of the demand for skill, as in Katz and Murphy (1992) and Card and Lemieux (2001). It is clear how this parameter is severely contaminated by selection bias, as is its trend. Furthermore, as we discuss below in section 3.2.3, composition changes can also affect the OLS parameter even in the absence of (fundamental) variation in parameters such as TT, ATE or TUT.

In section 3.2.1 we have described the evolution of prices of observed and unobserved skills in the economy. A natural question is: how do these price changes drive the evolution of inequality? To answer it we could, for example, follow the procedure of Juhn, Murphy and Pierce (1993) and decompose the increase in inequality in functions of observable and unobservable variables (or alternative procedures by DiNardo, Fortin and Lemieux, 1996, or Autor, Katz and Kearney, 2004). We adopt a similar idea but we account for selection, and we allow within group inequality to differ across skill groups (instead of just examining the residual of a wage regression pooling all schooling groups). In figure 8 we present the evolution of several measures of inequality in college and we break it into components due to observables and components due to unobservables. In the panels in the left hand column we account for selection while, for comparison, in the panels in the right hand column we do not, as is standard in the literature.[54] The panels in the first row plot changes in the 90-10 differential (difference between log wages in the 90th and 10th percentiles of the college distribution), those in the second row plot changes in the 90-50 differential and those in the third row plot changes in the 50-10 differential. In each panel we have three lines: one refers to the overall evolution of the inequality measure (circles), the other refers to the evolution we would observe in inequality if the distribution of U_1 were kept fixed at its 1992 values (triangles) and a third one

---

[53]Carneiro and Heckman (2002) show that this pattern can explain why IV estimates of the returns to schooling are usually larger than OLS estimates of this parameter.

[54]In the latter case we estimate college and high wage equations separately using ordinary least squares.

21

that plots the evolution of inequality if the distribution of $\mu_1(X, \beta_1)$ were kept at its 1992 values (squares). In the appendix B.4 we explain the exact procedure used.

Starting from figure 8, the first thing to notice is that the level of inequality in $f_{Y_1}(y_1)$ (left hand panels) is very different than the level of inequality in $f_{Y_1|S=1}(y_1|S=1)$ (right hand panels). For example, in 1992 the 90-10 differential in terms of $f_{Y_1}(y_1)$ is close to 1.5 while the corresponding measure in terms of $f_{Y_1|S=1}(y_1|S=1)$ is 1.26, a difference of more than 0.2. Even though the level of inequality (as measured by the 90-10 differential) is very different in the two top panels of figure 8, the trend in inequality is very similar. In both panels there is an increase of around 0.2 in this measure over the 1990s. However, when we look at the remaining panels of this figure we observe that this is not a general result: both the level and the trend in other measures of inequality are quite different whether we are looking to $f_{Y_1}(y_1)$ or $f_{Y_1|S=1}(y_1|S=1)$. The differences between the right and left panels of the figure exist for two reasons: (i) they concern the evolution of different objects ($f_{Y_1}(y_1)$ vs. $f_{Y_1|S=1}(y_1|S=1)$); (ii) the estimates of $\mu_1(X)$ on the right hand side panel are contaminated by selection bias.

It is clear from the three left panels of figure 8 that most of the increase in the evolution of counterfactual college inequality (which in principle accurately reflects changes in skill prices) is driven by changes in the distribution of unobservables (due, for example, to a rise in the price of unobserved ability, or abilities, which affects mostly the top of the wage distribution). Changes in the function $\mu_1(X)$ do not translate into significant changes in inequality at the top of the distribution (90-50 differential, presented in the second row of panels), but may have induced an increase in inequality at the bottom of the college wage distribution around 1996 (50-10 differential, shown in the third row of panels). When we do not account for selection (as is standard in the literature), we get very different answers. At the top of the wage distribution, the increase in "observed" inequality (not accounting for selection) not only is much smaller than the increase in counterfactual inequality, but its increase is attributable to both X and U_1, especially in the later years. At the bottom of the wage distribution observed inequality increases throughout the period and this is mainly due to X.[55] Given the results of table 3, it is not surprising that X plays a small role in the evolution of inequality, and it is also unlikely that skills associated with V play a substantial role. As emphasized in section 3.2.1, most of the variance of wages is driven by components of heterogeneity that are neither X nor V.

Figure 9 presents the same analysis for high school inequality. In this case, even though there are some differences between counterfactual and observed inequality, they are much smaller than in

---

[55]We could do similar types of decomposition using any other year as our base year. In results available on request we show that using 2000 as our base year instead of 1992 does not change our basic conclusions.

figure 8, probably because selection into high school does not play as important role in our paper as selection into college. The main determinant of the evolution of high school wage inequality is U_0. Movement in $\mu_0(X)$ does not translate into any significant change in high school wage inequality, although it plays some role in college inequality. This is probably due to the fact that AFQT, the main determinant of wages in our X vector, plays a more important role in college than in high school.

3.2.3 Composition Effects and Education Policy

The third exercise of our empirical section is an examination of the importance of composition effects. Unfortunately, since we follow a single cohort of individuals over time, there are no significant composition changes to examine in the raw data. Therefore, instead of looking directly to the data for evidence of composition effects, we adopt an alternative strategy. Using our estimates of the selection and outcome equations we simulate what would happen to inequality if college enrollment rates were different than the ones we observe, keeping prices fixed (partial equilibrium framework[56]). The difficulty of this exercise is to determine which individuals shift across schooling levels when the college enrollment rate changes. In general these individuals could come from anywhere in the potential college and high school wage distributions, but in principle those individuals at the margin between choosing to enroll in college or not will be the ones most likely to change their decision. In this paper we simulate changes in college enrollment by moving the intercept of the selection equation (equation (2)), and then we identify who are the individuals induced to go to college by this change (individuals at the margin). By looking at their potential high school and college wages we can determine how their movement affects the wage distribution. It is possible to imagine several other alternative ways to change the college enrollment rate in the economy (for example, through a tuition subsidy), and our conclusions about the effect of composition on inequality can certainly be affected by our choice. However, we feel this is a realistic approximation to a change in educational attainment. In this experiment, those individuals who are most likely to move first are just at the margin between attending college or not. As more and more people are induced to enter college they come further away from the margin. Therefore, even though we use one specific way to implement this simulation, the main patterns that we document will be present in any simulation that affects mostly individuals at the margin.

The mechanics of the simulation are simple: first we change the intercept of the schooling equation and we identify the distribution of (X, V) for individuals induced to enroll in college; second we

---

[56]See also Ferreira and Leite (2005).

generate the distribution of high school and college wages for this set of individuals; third, we compute how their exit from the high school sector affects the high school wage distribution and how their entry into the college sector affects the college wage distribution. The details of the simulation procedure are presented in Appendix B.5. The results are shown in figure 10, and they are based on our estimates for 1992 (we could have picked any other year as a base year to perform this exercise). We do not simulate settings where the average college enrollment rate is either very close to 0 or very close to 1. Instead, we restrict ourselves to cases where the average college enrollment rate is between 0.35 and 0.66 (the mean in the data is close to 0.5). An important concern with our simulations is that the model does not allow us to consider general equilibrium effects, and therefore does not allow skill prices to change as the quantities of different skills change in the economy. We hope that our simulation results will be more realistic if we restrict ourselves to a smaller (but still relatively large) range of values than the full possible range.

Figure 10 has six panels. In the top-left panel we show how the least squares measure of the college premium ($E(Y_1|S=1) - E(Y_0|S=0)$) changes as the probability of going to college increases. In the top right panel we show the evolution of the two components of the college premium: $E(Y_1|S=1)$ and $E(Y_0|S=0)$. The OLS estimate of this parameter rises sharply from 0.29 to 0.43 (an increase of 14%) as the proportion of individuals attending college decreases from 0.66 to 0.35 (a decrease of 31%). Roughly speaking, a 1 percent increase in college enrollment can induce a 0.5 percent decrease in this OLS estimate. The reason behind this is that the population of college graduates becomes more selective and of better quality as the cost of attending college increases, and therefore their average wages increase. Simultaneously, the population of high school graduates becomes of lower quality and average high school wages decrease (although this effect is very small in our data). As a result, the OLS estimate increases with the rise in the cost of attending college.

In the middle panels and the bottom left panel we look at trends in different quantiles of the college and high school wage distributions. In the bottom-right panel we examine changes in overall inequality (changes in the distribution of Y) as well as changes in within-skill-group inequality (changes in the distributions of college and high school wages, shown separately) by computing the difference between the 90th and the 10th percentiles of the distribution of Y, Y_1 conditional on $S=1$, and Y_0 conditional on $S=0$.

Median college wages decrease by 15% and median high school wages stay flat (an increase of about 1%) as the proportion of individuals with a college education increases from 0.35 to 0.66. The overall median wage increases by 5% as more individuals enter the college group. We observe the same behavior at the upper and lower tails of the distributions of high school, college and overall

24

wages. Furthermore, the observed (conditional on $S = 1$) 90-10 college wage differential increases by 7%, but the 90-10 high school wage differential stays essentially flat (a decrease of less than 1%). Again there is only a slight change in the distribution of high school wages. In summary, as the proportion of individuals attending college increases, between-group-inequality decreases, high school within-group-inequality stays flat and college wage inequality increases. Therefore, overall inequality can increase or decrease. In our sample, the overall 90-10 wage differential increases by 4% as college enrollment increases from 0.35 to 0.66.[57] In conclusion, relatively large increases in educational attainment can lead to relatively small increases in overall inequality.

At first glance it seems surprising to find large effects of composition on between group inequality but small effects on within group and overall inequalities. However, it is possible to reconcile these facts. In the presence of selection there can be large effects on parameters involving means at the same time that there are small effects on dispersion parameters. This will happen if the amount of heterogeneity on which individuals select does not explain a lot of the dispersion in wages. Indeed, results already described in table 3 confirm that only a small amount of total variance is explained by selection and therefore it is quite conceivable that selection and composition changes have little effect on the dispersion parameters. In spite of this, as shown in section 3.2.2, selection bias substantially changes the description of within group wage inequality. When we examine selection bias we compare the observed data to very extreme counterfactuals: we compare $f_{Y_1|S=1}(y_1|S=1)$ to $f_{Y_1}(y_1)$ and $f_{Y_0|S=0}(y_0|S=0)$ to $f_{Y_0}(y_0)$. This is analogous to simulating very large changes in composition, shifting everyone to one schooling level of the other. Our analysis of selection bias shows that, in spite of the results of table 3, such extreme movements in composition can have an effect on wage dispersion.

The analysis of figure 10 suggests that composition changes may play some role in explaining changes in inequality, especially changes in the least squares measure of the college premium (although so far we experimented with larger changes in educational attainment than the ones observed in the data, say, over a decade). For example, over the 1980s, a period of sustained increase in inequality, using the Census we observe that the percentage of white males aged 25-65 with some college or more increased from 41% (1980) to 55% (1990). How did this change in college enrolment affect the rise in inequality? Using our estimates of the selection model and taking as the base year 1992, we simulate a change in college enrollment from 41% to 55%. The numbers we obtain from the simulation are not directly comparable with the ones we obtain from the Census, since the age

---

[57]The relationship between college enrollment and overall inequality does not even need to be monotone. Given that between group and within group inequality move in opposite directions, it is simple to design very reasonable examples where a non-monotonic relationship arises.

group with which we estimated the selection model is quite restricted and our base year is 1992, but this exercise is still instructive. Table 4 shows the result of our experiment. The first columns are computed from the US Census, and the second two columns correspond to the simulations. The OLS estimate of the return to schooling decreases from 40% to 33%. This indicates that, if anything, the measured return to schooling in the 1980s would have increased by 7% more in the absence of composition changes. Furthermore, if the 1980s description of inequality is affected by selection bias in the same way as the 1990s description of inequality (as seen in section 3.2.2), there is a further understatement of the trend in the average return to college in the 1980s.

The changes in within group inequality and overall inequality, as measured by the 90-10 percentile wage differential, are quite smaller. In college, this measure increases from 1.30 to 1.34, an increase of 4%, while in the Census the 90-10 differential in college increases by 11%. In high school, the 90-10 percentile wage differential decreases from 1.13 to 1.12, a decrease of 1%, while in the Census the 90-10 differential in high school increases by 9%.[58] Overall there is a very small increase in inequality, which increases from 1.28 to 1.29.[59] Nevertheless, the effect on the standard measure of between group inequality is quite substantial (the OLS estimate of the return to college decreases by 7%), indicating that selection and composition effects do play an important role. As mentioned above, this can have important consequences for studies of the evolution of inequality and of the returns to schooling, such as Katz and Murphy (1992), Card and Lemieux (2003) or Gosling, Machin and Meghir (2000).

The results in this section also have immediate consequences for the study of the effects of education policy on wage inequality in a partial equilibrium framework. Policies that increase college attendance lead to substantial reductions in the average "quality" of college graduates and may lead to some increase in the average "quality" of high school graduates. As a consequence, the difference between college and high school average wages is likely to decline. These effects are substantial even for increases in college enrollment of about 10-15% (comparable to the increase in educational attainment observed in the 1980s). At the same time there are much smaller changes in within group and overall inequalities. This suggests that education policy can have large effects on the OLS measure of the return to schooling, but much smaller effects on the overall level of inequality. The reason is that, even though there is selection into college, the variables on which individuals select explain a small fraction of the total variance of wages. Furthermore, the difference

---

[58]Therefore we "explain" roughly 30% of the increase in within group inequality in college with composition changes. But we explain none of the increase in within group inequality in high school because in our simulation, changes in composition give rises to a small decrease in within group inequality while the same measure increased in the Census data.

[59]A larger change in composition could lead to a substantial change in inequality. In the extreme, given our 1992 estimates, overall inequality would increase from 1.26 to 1.49 if everyone were forced to enroll in college.

in average wages of college and high school graduates is also a very small fraction of the overall dispersion in wages, and wages are more disperse in college than in high school.[60]

4 Conclusion

Most of the literature on inequality ignores that individuals are heterogeneous and self-select into different levels of schooling. Our paper is an attempt to account for selection into schooling in the empirical study of inequality. We focus on a sample of white males from the NLSY, and examine the evolution of their wages during the 1990s. This is neither the sample that is commonly used in studies of this type (such as the CPS, or the Census), nor the time period with greater changes in inequality (the 1980s). We pick this sample for a reason: its richness of individual information allows to model the process of selection into schooling in a way that is almost impossible in other data. Furthermore, this is a nationally representative sample of males, with a reasonable sample size, and the wage distributions are very similar for the CPS and the NLSY during the 1990s (for the same cohorts of white males).

We estimate a semiparametric selection model with two levels of schooling (high school and college) using five years of data (1992, 1994, 1996, 1998 and 2000) and use it for three different exercises. First, we use them to understand the main patterns of sorting of individuals into different levels of schooling. We find that individuals sort into the level of schooling where they have absolute and comparative advantage. Second, we use them to analyze the evolution of inequality and of its determinants in our sample during the 1990s, purging our estimates of selection bias. We find that selection bias causes us to understate the growth in different measures of the average return to schooling. Furthermore, it leads us to understate the increase in inequality at the top of the college wage distribution and to understate it at the bottom of the college wage distribution. Third, we use them to simulate changes in schooling attainment and examine their effects on the wage structure. In our data increases in educational attainment lead to reductions in between group inequality but do not have substantial effects on within group inequality, unless the change in educational attainment is very large.

Some of the effects we emphasize in this paper are also present in some analyses of larger samples, in particular the Census. Juhn, Kim and Vella (2005) find evidence of cohort quality effects systematically related to the educational attainment of different cohorts, but argue these can only explain a small fraction of recent fluctuations in the college premium. Using a different approach,

---

[60]This also suggests that even in the absence of selection, it may well be that if we induce a large fraction of the population to enroll in college we will not have a large reduction in inequality (or poverty).

Carneiro and Lee (2005) also find suggestive evidence of composition effects using Census data from 1960 to 2000, by linking differences in college and high school wages across cohorts to educational attainment across cohorts, accounting for equilibrium effects of college enrollment on skill prices (as emphasized by, for example, Card and Lemieux, 2001). Furthermore, using data from the International Adult Literacy Survey for the US, they also show that (after controlling for age effects) cohorts with relatively large college enrollment rates in the sample tend to have lower average literacy wages among college graduates, indicating a decline in the average quality of the college graduate in these cohorts.

In summary, the literature on the returns to schooling emphasizes that heterogeneity and self-selection are important features of labor markets. In this paper we show that selection into schooling seems to be an important factor to take into account in studies of wage dispersion. It can substantially change our perception of the forces that shape the evolution of wage inequality.

Appendix

A   Description of the Data

We restrict the NLSY sample to white males. We define individuals in the high school group as individuals having up to a high school degree, or having completed up to 12 grades and never reporting college attendance. We define participation in college as having ever gone to college or having complete more than 12 grades in school. GED recipients that do not have a high school degree, who have less than 12 years of schooling completed and who never reported college attendance are excluded from the sample. The wage variables we use are deflated (to 1983) non-missing hourly wages from 1990 to 2002. We use these to construct 5 year averages which we use in the analysis. We delete all wage observations that are below \$1 and above \$100. Experience is actual work experience in weeks accumulated from 1979 to the year of interest (annual weeks worked are imputed to be zero if they are missing in any given year). The remaining variables that we include in the X and Z vectors are number of siblings, father's years of schooling, mother's years of schooling, schooling corrected AFQT, average deflated (to 1993) tuition of the colleges in the county the individual lives in at 19, distance to the nearest college at age 14 and local unemployment rate in state of residence in 1979. For the construction of the tuition variable see Cameron and Heckman (2001). Distance to college is constructed by matching college location data in HEGIS with county of residence in NLSY. For a description of the NLSY sample see BLS (2001). The NLSY79 has an oversample of poor whites which we exclude from this analysis. We also exclude the military sample. To remove the effect of schooling on AFQT we implement the same procedure as in Carneiro, Heckman and Vytlacil (2003) (based on Hansen, Heckman and Mullen, 2004).

B   Details of Estimation Procedure

B.1   Estimating $E[U_1|V = v]$ and $E[U_0|V = v]$

This section gives a detailed description of nonparametric estimators of $E[U_1|V = v]$ and $E[U_0|V = v]$. First consider local polynomial estimation of $E[U_1|V = v]$. In general, use of higher order polynomials may reduce the bias but increase the variance by introducing more parameters. Fan and Gijbels (1996) suggest that the order π of polynomial be equal to $\pi = \mu + 1$, where μ is the order of the derivative of a function of object. That is, Fan and Gijbels (1996) recommend a local linear estimator for fitting a function and a local quadratic estimator for fitting a first-order derivative. Following suggestions of Fan and Gijbels (1996), $E(U_1|P = v, S = 1)$ is estimated by a local linear estimator using observations with $S = 1$ and $\partial E(U_1|P = v, S = 1)/\partial p$ is estimated by a local quadratic estimator.

To be more specific, let $\{(\hat{U}_{1i}, \hat{P}_i, S_i) : i = 1, \ldots, n\}$ denote observations of estimated U_1 and P along with S, where $\hat{U}_{1i} = Y_i - \mu_1(X_i, \hat{\beta}_1)$ for $i = 1, \ldots, n$. The local linear estimator $\hat{E}(U_1|P = v, S = 1)$ is obtained by solving the problem

$$\min_{c_0, c_1} \sum_{i=1}^{n} S_i \left[\hat{U}_{1i} - c_0 - c_1(\hat{P}_i - v) \right]^2 K\left(\frac{\hat{P}_i - v}{h_{n1}} \right),$$

where $K(\cdot)$ is a kernel function and h_{n1} is a bandwidth. The resulting value of c_0 is the local linear estimator of $E(U_1|P = v, S = 1)$. Similarly, the local quadratic estimator $\hat{\partial} E(U_1|P = v, S = 1)/\partial p$ is obtained by solving the problem

$$\min_{c_0, c_1, c_2} \sum_{i=1}^{n} S_i \left[\hat{U}_{1i} - c_0 - c_1(\hat{P}_i - v) - c_2(\hat{P}_i - v)^2 \right]^2 K\left(\frac{\hat{P}_i - v}{h_{n2}} \right),$$

where h_{n2} is a bandwidth that can be different from h_{n1}. The resulting value of c_1 is the local quadratic estimator of $\partial E(U_1|P = v, S = 1)/\partial p$. Then the estimator of $E[U_1|V = v]$ is given by

$$\hat{E}[U_1|V = v] = v \frac{\hat{\partial}}{\partial p} E(U_1|P = v, S = 1) + \hat{E}(U_1|P = v, S = 1). \tag{9}$$

Similarly, the estimator of $E[U_0|V = v]$ can be obtained by replacing unknown functions in the right hand side of (7) with their nonparametric estimators.

In our empirical work, we first need to estimate β_1 and β_2. They are estimated by a Robinson (1988)-type estimator with a bandwidth of 0.10. The data are trimmed by 10% of the observations with the smallest density estimates of the estimated P. The estimation results do not change much as we use alternative bandwidths (0.05 and 0.20) or an alternative trimming (5%).

Estimating $E(U_1|P = v, S = 1)$ and its derivative requires choices of two bandwidths h_{n1} and h_{n2}. A reasonable data-driven bandwidth selection rule is important to carry out nonparametric estimation. We carry out some initial search for bandwidths using a method called residual squares criterion (RSC) proposed in Fan and Gijbels (1996, Section 4.5). After experimenting different bandwidths around RSC-chosen bandwidths, we finally choose $h_{n1} = 0.35$ and $h_{n2} = 1.25h_{n1}$ for estimating both $E(Y_1|X, V)$ and $E(Y_0|X, V)$ for all the years. Varying the value of h_{n1} from 0.2 to 0.5 does not make any important changes in the estimation results. Throughout the paper, we use

the standard normal density function as the kernel function K. Bootstrap confidence intervals shown in figures 2 and 3 are obtained by the bootstrap percentile method with 499 bootstrap replications.

B.2 Estimating $f(u_1|v)$ and $f(u_0|v)$

This section describes nonparametric estimators of $f(u_1|v)$ and $f(u_0|v)$. As in (6) and (7), an application of Lemma (1) yields the following relationships

$$f_{U_1|V}(u_1|v) = f_{U_1|P,S=1}(u_1|v, S = 1) + v\frac{\partial}{\partial p}f_{U_1|P,S=1}(u_1|v, S = 1) \quad \text{and} \tag{10}$$

$$f_{U_0|V}(u_0|v) = f_{U_0|P,S=0}(u_0|v, S = 0) - (1 - v)\frac{\partial}{\partial p}f_{U_0|P,S=0}(u_0|v, S = 0). \tag{11}$$

Sample analogs of the right-hand sides of equations (10) and (11) can be obtained by some suitable nonparametric estimators.

We only discuss estimation of $f(u_1|v)$ in detail, since estimation of $f(u_0|v)$ is similar. To develop an estimator of $f(u_1|v)$ using the equation (10), it is necessary to estimate $f_{U_1|P,S=1}(u_1|p, S = 1)$ and its derivative with respect to p. Specifically, the estimator of $f(u_1|v)$ can be obtained by

$$\hat{f}(u_1|v) = v\frac{\partial}{\partial p}f_{U_1|P,S=1}(u_1|v, S = 1) + \hat{f}_{U_1|P,S=1}(u_1|v, S = 1), \tag{12}$$

where $\hat{f}_{U_1|P,S=1}(u_1|v, S = 1)$ and $\partial\hat{f}_{U_1|P,S=1}(u_1|v, S = 1)/\partial p$ are defined below.

In order to compute $\hat{f}_{U_1|P,S=1}(u_1|v, S = 1)$ and $\partial\hat{f}_{U_1|P,S=1}(u_1|v, S = 1)/\partial p$ in (12), we begin with estimated data $\{(\hat{U}_{1i}, \hat{P}_i) : i = 1, \ldots, n, S_i = 1\}$, where $\hat{U}_{1i} = Y_i - \mu_1(X_i, \hat{\beta}_1)$. One could estimate the conditional density of U_1 given P and its derivative by estimating the joint and marginal densities using the standard kernel density estimators, and then taking the ratio between them to estimate the conditional density, and finally computing a derivative of the conditional density. This indirect method would yield consistent estimators but it is quite cumbersome. Instead we use a direct method of Fan, Yao, and Tong (1996), who develop local polynomial estimators of the conditional density function and its derivative. To motivate the estimators of Fan, Yao, and Tong (1996), notice that, as $\delta_n \to 0$,

$$E\left[\delta_n^{-1}K\left(\frac{U_1 - u_1}{\delta_n}\right)\bigg| P = v, S = 1\right] \approx f_{U_1|P,S=1}(u_1|v, S = 1)$$

$$\approx f_{U_1|P,S=1}(u_1|v_0, S = 1) + \frac{\partial}{\partial p}f_{U_1|P,S=1}(u_1|v_0, S = 1)(v - v_0)$$

for any v in a neighborhood of v_0, where K is a nonnegative density function and δ_n is a bandwidth. This suggests that the local linear estimator of $f_{U_1|P,S=1}(u_1|v, S = 1)$ can be defined as $\hat{f}_{U_1|P,S=1}(u_1|v, S = 1) \equiv \hat{c}_0$, where $(\hat{c}_0, \hat{c}_1)$ solves the problem

$$\min_{c_0,c_1}\sum_{i=1}^{n} S_i\left[\delta_n^{-1}K\left(\frac{\hat{U}_{1i} - u_1}{\delta_n}\right) - c_0 - c_1(\hat{P}_i - v)\right]^2 K\left(\frac{\hat{P}_i - v}{h_{n1}}\right), \tag{13}$$

and the local quadratic estimator of $\partial f_{U_1|P,S=1}(u_1|v, S = 1)/\partial p$ can be defined as $\partial\hat{f}_{U_1|P,S=1}(u_1|v, S = 1)/\partial p \equiv \hat{c}_1$, where $(\hat{c}_0, \hat{c}_1, \hat{c}_2)$ solves the problem

$$\min_{c_0,c_1,c_2}\sum_{i=1}^{n} S_i\left[\delta_n^{-1}K\left(\frac{\hat{U}_{1i} - u_1}{\delta_n}\right) - c_0 - c_1(\hat{P}_i - v) - c_2(\hat{P}_i - v)^2\right]^2 K\left(\frac{\hat{P}_i - v}{h_{n2}}\right). \tag{14}$$

The estimator defined in (12) is an unrestricted estimator. Thus, it can be negative for a given finite sample, although it is a consistent estimator of $f(u_1|v)$ under certain regularity conditions. To ensure that the estimator is positive in finite samples, we consider a trimmed version of (12):

$$\hat{f}_{pdf}(u_1|v) = \max[\varepsilon, \hat{f}(u_1|v)],$$

where ε is a fixed, very small positive number.

Now we describe estimators of $F(u_1|v)$ and $F(u_0|v)$. Again we only discuss estimation of $F(u_1|v)$. To develop an estimator of $F(u_1|v)$, note that

$$F(u_1|v) = F_{U_1|V}(\underline{u}_1|v) + \int_{\underline{u}_1}^{u_1} f_{U_1|V}(u|v)du, \tag{15}$$

for any fixed constant $\underline{u}_1 < u_1$. We estimate $F(u_1|v)$ by replacing $F_{U_1|V}(\underline{u}_1|v)$ and $f_{U_1|V}(u|v)$ in (15) with their sample analogs. More specifically, the estimator of $F_{U_1|V}(\underline{u}_1|v)$ is defined as

$$\hat{F}_{U_1|V}^{cdf}(\underline{u}_1|v) = \max[0, \hat{F}_{U_1|V}(\underline{u}_1|v)],$$

where

$$\hat{F}_{U_1|V}(\underline{u}_1|v) = v\frac{\hat{\partial}}{\partial p}F_{U_1|P,S=1}(\underline{u}_1|v, S=1) + \hat{F}_{U_1|P,S=1}(\underline{u}_1|v, S=1), \tag{16}$$

and $\hat{F}_{U_1|P,S=1}(\underline{u}_1|v, S=1)$ and $\hat{\partial}F_{U_1|P,S=1}(\underline{u}_1|v, S=1)/\partial p$, respectively, are local linear and quadratic estimators that solve the problems similar to those in (13) and (14) with $\delta_n^{-1}K\left((\hat{U}_{1i} - u_1)/\delta_n\right)$ replaced by $1(\hat{U}_{1i} \leq \underline{u}_1)$. Then our estimator of $F(u_1|v)$ is defined as

$$\hat{F}_{cdf}(u_1|v) = \min\left[1, \hat{F}_{U_1|V}^{cdf}(\underline{u}_1|v) + \int_{\underline{u}_1}^{u_1} \hat{f}_{pdf}(u|v)du\right].$$

Notice that by construction, our estimator is a strictly increasing, continuous function of u_1 (for $u_1 > \underline{u}_1$) and is restricted to be between 0 and 1. In other words, our estimator is a distribution function for a given finite sample. One could use an unrestricted estimator (16), which is not necessarily a distribution function in finite samples.

To estimate these conditional PDF's and CDF's, we adopt the same bandwidths h_{n1} and h_{n2} that are used to estimate the corresponding conditional means. The bandwidth δ_n is chosen by Silverman's normal reference rule (Silverman, 1986, p.45). These choices of bandwidths are arbitrary, but the estimation results are not very sensitive to the choices of the bandwidths.

Notice that as a by-product of estimating $\hat{F}_{cdf}(u_1|v)$, we obtain an estimator of the τ-th quantile of U_1 conditional on $V = v$ for any $\tau \in (0,1)$, which is denoted by $Q_{U_1|V}(\tau|v)$. Simply, the estimator is given by

$$\hat{Q}_{U_1|V}(\tau|v) = \hat{F}_{cdf}^{-1}(\tau|v),$$

where the right-hand side is unique for a given finite sample provided that $\underline{u}_1$ is sufficiently small, since $\hat{F}_{cdf}(u_1|v)$ is a strictly increasing function when $u_1 > \underline{u}_1$. Furthermore, under the assumption that U_1 and V are independent of X and Z, the τ-th quantile of Y_1 conditional on $X = x$ and $V = v$ can be estimated by

$$\hat{Q}_{Y_1|X,V}(\tau|x,v) = \mu_1(x, \hat{\beta}_1) + \hat{Q}_{U_1|V}(\tau|v).$$

Therefore, we can also obtain estimators of quantile treatment effects, which are defined as

$$\hat{Q}_{Y_1|X,V}(\tau|x,v) - \hat{Q}_{Y_0|X,V}(\tau|x,v).$$

In the recent econometrics literature, quantile treatment effects have been considered by Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005), Chesher (2003), and Imbens and Newey (2003) among others.

B.3 Obtaining Sample Analogs of (8)

Estimators of $E[Y_j]$, $E[Y_j|S = 1]$, and $E[Y_j|S = 0]$ are obtained by

$$E[Y_j] = n^{-1} \sum_{i=1}^{n} \int_0^1 \hat{E}[Y_j|X = X_i, V = v] \, dv,$$

$$E[Y_j|S = 1] = n^{-1} \sum_{i=1}^{n} \int_0^1 \hat{E}[Y_j|X = X_i, V = v] \frac{1 - \hat{F}_{P|X}(v|X_i)}{\hat{\Pr}(S = 1)} \, dv, \tag{17}$$

and

$$E[Y_j|S = 0] = n^{-1} \sum_{i=1}^{n} \int_0^1 \hat{E}[Y_j|X = x, V = v] \frac{\hat{F}_{P|X}(v|X_i)}{\hat{\Pr}(S = 0)} \, dv,$$

where $\hat{E}[Y_j|X = x, V = v] = \mu_j(x, \hat{\beta}_j) + \hat{E}[U_j|V = v]$ is defined in the main text, $\hat{F}_{P|X}(v|x)$ is a nonparametric estimator of $F_{P|X}(v|x)$, and $\hat{\Pr}(S = j)$ is the sample proportion of $S = j$ for $j = 0, 1$. The integration with respect to v can be evaluated numerically.

B.4 Estimating Counterfactual Changes in Inequality as a Function of Observables and Unobservables

This subsection describes how we obtain results shown in Figure 8. The results in Figure 9 are obtained similarly.

Figure 8 consists of 6 panels. The left-hand panels exhibit 90-10, 90-50 and 50-10 differences of the counterfactual distribution of Y_1. In these figures, solid lines with circles are obtained from estimates of the counterfactual distributions of Y_1 over different years; closely dotted lines with squares are computed from estimates of the counterfactual distributions of Y_1 by on one hand, varying estimates of $\mu_1(x, \beta_1)$ over different years and on the other hand, fixing f_{U_1} at the 1992 estimate; dotted lines with triangles are from estimates of the counterfactual distributions of Y_1 by on one hand, varying estimates f_{U_1} over different years and on the other hand, fixing $\mu_1(x, \beta_1)$ at the 1992 estimate. To carry out this procedure, it is necessary to impose location normalization on U_1. We achieve this by setting $E[U_1] = 0$ for all the years.

The right-hand panels show 90-10, 90-50 and 50-10 differences of the observed distributions of Y_1 conditional $S = 1$. To obtain these figures, we run an OLS regression of Y_1 on X using only $S = 1$ to obtain estimates of $\mu_1(x, \beta_1)$ and U_1. Solid lines with circles are obtained directly from observed data on college wages over different years; closely dotted lines with squares are computed from OLS estimates of $\mu_1(x, \beta_1)$ over different years and the 1992 OLS residuals; the dotted lines with triangles are from the 1992 OLS estimate of $\mu_1(x, \beta_1)$ and OLS residuals over different years.

B.5 Simulating Wage distributions of Different Compositions of Education Groups

This subsection describes how we carry out simulations whose results are shown in Figure 10. Suppose there is a policy that shifts the distribution of P in the population from $F_{P|X}(p|x)$ to $F^*_{P|X}(p|x)$, but has no effect on $f(y_0|x,v)$ nor $f(y_1|x,v)$ (for example, tuition subsidies). In view of (8), the post-policy distributions of college and high school wages are

$$f(y_1|S=1) = \int\int_0^1 f(y_1|x,v)\frac{1-F^*_{P|X}(v|x)}{\Pr^*(S=1)}f_X(x)\,dv\,dx,$$

and

$$f(y_0|S=0) = \int\int_0^1 f(y_0|x,v)\frac{F^*_{P|X}(v|x)}{\Pr^*(S=0)}f_X(x)\,dv\,dx. \qquad (18)$$

Thus, in order to simulate wage distributions of different compositions of education groups, we need to compute only $F^*_{P|X}(p|x)$ and $\Pr^*(S=1)$. Recall that in our empirical work, $P = L(\mu_S(Z))$, where L is the CDF of the logistic distribution, We simulate changes in college enrollment rates simply by varying the intercept of μ_S. Then $F^*_{P|X}(p|x)$ can be estimated by a nonparametric kernel regression of $1\{L(\hat{\mu}_S(Z)+c_*)\le p\}$ on X with an estimated choice index $\hat{\mu}$ and $\Pr^*(S=1)$ can be estimated by a sample average of $L(\hat{\mu}_S(Z)+c_*)$, where we vary c_* from a small number to a large number to simulate a range of college enrollment rates. Finally the results shown in Figure 10 can be obtained by sample analogs of (18).

C Goodness-of-fit of the Model Specification and Tests for Selection

This section report an informal goodness-of-fit check of our model specification and results of simple tests for selection. First, we provide evidence on the goodness-of-fit of our model specification. Figure 12 compares estimates of $E(Y_1|S=1)$, $E(Y_0|S=0)$, Quantile $(Y_1|S=1)$ and Quantile $(Y_0|S=0)$ from the model with the corresponding quantities in the data, for all the years of our analysis. Overall, our model fits the data relatively well. Even though the estimation of these objects involves a multiple step procedure, our estimates of the means of high school and college wages and the data match almost perfectly, although the fit is not as good in the lower tail of the college and high school wage distributions. This gives us some confidence in the specification of the model.

We next report results of simple tests for selection, summarized in table 6. We carry out tests for selection for both college and high school wages. Four different null hypotheses are tested: (1) no selection on unobservables, (2) no selection on family background and unobservables, (3) no selection on AFQT and unobservables, and (4) no selection on AFQT, family background and unobservables. We describe a test for the first null hypothesis in detail. Other tests are similar. Our test is based on the fact that $E[(E[Y_j|X]-E[Y_j|X,Z])^2]\ge 0$ with equality holding only if $E[Y_j|X]=E[Y_j|X,Z]$ almost surely, in which case we have no selection on unobservables but selection on only observables. Although there is a literature on testing this type of hypothesis against a general nonparametric alternative (see, for example, Fan and Li 1996), we decide to carry out a simpler test against a

parametric but flexible alternative. Specifically, our simple test for no selection on unobservables is a Hausman-type bootstrap test, which is based on the mean square difference between fitted values of the restricted and unrestricted parametric models. The restricted model is a linear regression model with X whose specification is already described in section 3.1. The unrestricted model is a linear regression model with the same X above along with the following additional regressors: linear, quadratic, and interaction terms of instruments (tuition, distance, unemployment in 1979) and interactions between instruments and AFQT and family background variables (number of siblings, mother's education, and father's education). For tests for the null hypotheses (2)-(4), the restricted models are linear regression models with similar specifications except that corresponding observables are omitted from the specifications and the unrestricted models are identical to one used for testing the null hypothesis (1). Each p-value reported in table 6 is obtained (without Studentization) by wild bootstrap (e.g. see Wu, 1986) imposing the null hypothesis with 999 bootstrap replications. Finally, for comparison, the likelihood ratio (LR) test is carried out via the maximum likelihood estimation using the normal selection model.

It can be seen that for college wages, the null hypothesis of no selection on unobservables is rejected at 10% level for all the years and rejected at 5% level for three years, and for high school wages we fail to reject no selection on unobservables at any conventional levels ($p > 0.1$) for all the years. These results conform to those of the LR tests using the normal selection model. It can also be seen that (1) AFQT is an important determinants of selection for both college and high school wages, (2) family background variables matter only for selection in terms of college wages, and (3) the null hypothesis of no selection on AFQT, family background and unobservables is rejected at all conventional levels ($p < 0.01$) almost for all college and high school wages (except for high school wages in 2000 with p-value of 0.044). These results suggest that selection on unobservables may exist for standard analyses of wage inequality because a measure of cognitive ability and/or family background variables are usually absent in standard analyses.

We end this section by mentioning some limitation of our simple tests. Evidence against selection on unobservables for high school wages is by no means conclusive. One caveat of our tests is that they may have a low power against an alternative that is different from the one we specify above (or our test may even not be consistent). To see how sensitive our bootstrap tests are against different alternatives, we carry out some Monte Carlo experiments in which we use same S, X and Z as in 1992 and simulate Y using our estimates (details are available on request). We find that indeed our test for no selection on unobservables (the null hypothesis (1)) is much less powerful against alternatives are not nested in our specification. However, when the null hypothesis is correct, then there is virtually no difference between the true and nominal probabilities of rejecting the null hypothesis. The main purpose of this paper is to analyze wage inequality allowing for selection rather than to prove that there is no selection. Therefore, we are on the conservative side and consequently we opt not to develop a more powerful test. One main concern with a test against a general nonparametric alternative is the curse of dimensionality. It would be interesting to develop an omnibus test that is free from the curse of dimensionality in the setup of selection models, but this is beyond the scope of this paper.

References

Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97: 284-292.

Abadie, A., J. D. Angrist, and G. W. Imbens (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70: 91-117.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000), "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67: 499-527.

Autor, D., L. Katz and A. Krueger (1998), "Computing Inequality: Have Computers Changed the Labor Market?", *Quarterly Journal of Economics*, 113 (4), 1169-1214.

Autor, D., L. Katz and M. Kearney (2004), "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists", working paper, MIT.

Björklund, A. and R. Moffitt (1987), "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69:42-49.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2004), "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition using Bounds," IFS working paper, W04/25.

Blundell, R., H. Reed and T. Stoker (2003), "Interpreting Aggregate Wage Growth: The Role of Labour Market Participation", *American Economic Review*, 93:1114-1131.

Browning, M., L. Hansen and J. Heckman (1999), "Micro Data and General Equilibrium Models", in J. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, Amsterdam:Elsevier.

Card, D. (1993), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling", NBER wp 4483.

Card, D. (1999), "The Causal Effect of Education on Earnings," Orley Ashenfelter and David Card, (editors), Vol. 3A, *Handbook of Labor Economics,* Amsterdam: North-Holland.

Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5): 1127-60.

Card, D. and T. Lemieux (2001), "Can Falling Supply Explain the Rising Return to College For Younger Men? A Cohort Based Analysis," *Quarterly Journal of Economics* 116: 705-46.

Cameron, S. and J. Heckman (2001). "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109:455-499.

Cameron, S. and C. Taber (2004), "Estimation of Educational Borrowing Constraints Using Returns to Schooling", *Journal of Political Economy*, part 1, 112(1): 132-82.

Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post-secondary Schooling," *Economic Journal* 112(482): 705-34.

Carneiro, P., J. Heckman and E. Vytlacil (2003), "Understanding what Instrumental Variables estimate: estimating average and marginal returns to schooling", working paper, The University of Chicago.

Carneiro, P., K. Hansen and J. Heckman (2003), "Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice," *International Economic Review*, 44(2): 361-422.

Carneiro, P. and S. Lee (2004), "Comparative Advantage and Schooling", working paper, University College London.

Carneiro, P. and S. Lee (2005), "Changes in Educational Attainment and Inequality: Sorting and Equilibrium Effects", working paper, University College London.

Chay, K. Y. and D. S. Lee (2000), "Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black-White Wage Differentials," *Journal of Econometrics*, 99: 1-38.

Chandra, A. (2003), "Is the Convergence of the Racial Wage Gap Illusory?", NBER working paper 9476.

Chernozhukov, V. and C. Hansen (2005), "An IV model of Quantile Treatment Effects," *Econometrica*, 73(1): 245-61.

Chesher, A. D. (2003), "Identification in Nonseparable Models," *Econometrica*, 71: 1405-1441.

Cunha, F., J. Heckman and S. Navarro (2005), "Separating Heterogeneity From Uncertainty In Income and Human Capital Dynamics", working paper, The University of Chicago.

Das, M., W. K. Newey and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economics Studies*, 70: 33-58.

DiNardo, Fortin and Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica*, 64 (5), 1001-1044.

Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications* (London: Chapman & Hall).

Fan, J., Q. Yao, and H. Tong (1996), "Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems," *Biometrika*: 83:189-206.

Fan, Y. and Q. Li (1996), "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64(4): 865-90.

Ferreira, F. and P. Leite (2005), Educational Expansion and Income Distribution: a Micro-Simulation for Ceara, in A. Shorrocks and R. Hoeven, *Growth, Inequality and Poverty: Prospects for Pro-poor Economic Development*, Oxford University Press.

Gosling, A., S. Machin, and C. Meghir (2000), T"he Changing Distribution of Male Wages in the U.K.," *Review of Economic Studies*, 67(4): 635-66.

Gould, E. D. (2002), "Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States," *Journal of Labor Economics*, 20(1): 105-47.

Hansen, K., K. Mullen and J. Heckman (2004), "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*. 121(1-2): 39-98.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.

Heckman, J., L. Lochner and C. Taber (1998), "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, 1(1): 1-58.

Heckman, J. and G. Sedlacek (1985), "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *Journal of Political Economy*, Vol. 93, No. 6., pp. 1077-1125.

Heckman, J. and J. Smith (1998), "Evaluating the Welfare State," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*, Econometric Monograph Series, ed. by S. Strom, Cambridge, UK: Cambridge University Press.

Heckman, J., J. Smith, and N. Clements (1997), "Making the Most out of Program Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts," *Review of Economic Studies* 64, 487-535.

Heckman, J. and P. Todd (2000), "Understanding the Contribution of Legislation, Social Activism, Markets, and Choice to the Economic Progress of African Americans in the Twentieth Century", chapter prepared for the American Bar Foundation.

Heckman, J. J. and E. Vytlacil (1999), Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 96, 4730.4734.

Heckman, J. and E. Vytlacil (2001), "Local Instrumental Variables," in C. Hsiao, K. Morimune, and J. Powells, (eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, (Cambridge: Cambridge University Press, 2001), 1-46.

Heckman, J. and E. Vytlacil (2005), "Structural Equations, Treatment, Effects and Econometric Policy Evaluation," *Econometrica*, 73(3):669-738.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2):467-475.

Imbens, G. and W. Newey (2003), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity", working paper, MIT.

Imbens, G. and D. Rubin (1997), "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economics Studies*, 64: 555-574.

Juhn, C., D. -L. Kim, and F. Vella (2005), "The Expansion of College Education in the United States: Is There Evidence of Declining Cohort Quality?," *Economic Inquiry*, 43(2): 303-15.

Juhn, C., K. Murphy and B. Pierce (1993), "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, 101(3):410-442.

Kane, T. and C. Rouse (1995), "Labor-Market Returns to Two- and Four-Year College", *American Economic Review*, 85(3):600-614.

Katz, L. and D. Autor (1999), "Changes in the Wage Structure and Earnings Inequality", in O. Ashenfelter and D. Card, eds, *Handbook of Labor Economics*, volume 3, North-Holland.

Katz, L. and K. Murphy (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *Quarterly Journal of Economics*, Vol. 107, No. 1. (Feb., 1992), pp. 35-78.

Kling, J. (2001), "Interpreting Instrumental Variables Estimates of the Returns to Schooling", *Journal of Business and Economic Statistics*, 19(3), 358-364.

Lemieux, T. (2004), "Increased Residual Wage Inequality: Composition Effects, Noisy Data or Rising Demand for Skill", Center for Labor Economics Working Paper No 69, UC Berkeley

National Center for Education Statistics (2003), *Digest of Education Statistics*, 2002, US Department of Education.

Neal, D. (2004), "The Measured Black-White Wage Gap among Women Is Too Small", *Journal of Political Economy*, Part 2, 112(1): S1-28.

Robinson, P. M. (1988), "Root- N-Consistent Semiparametric Regression," *Econometrica*, 56(4): 931-54.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Taber, C. R. (2001), "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *Review of Economic Studies*, 68(3): 665-91.

Welch, F. (1999), "In Defense of Inequality", *American Economic Review*, 89(2), 1-17.

Willis, R. and S. Rosen (1979), "Education and Self-Selection," *Journal of Political Economy*, 87(5):Pt2:S7-36.

Wu, C. F. J. (1986), Jacknife, Boostrap, and Other Resampling Methods in Regression Analysis (with discussion), *Annals of Statistics*, 14: 1261-1295.

Table 1: Summary Statistics of Data

Variable	Year 1992 Mean	SD	Year 1994 Mean	SD	Year 1996 Mean	SD	Year 1998 Mean	SD	Year 2000 Mean	SD
College ($S = 1$)	$n = 822$		$n = 806$		$n = 779$		$n = 773$		$n = 750$	
Log Wage	2.72	0.56	2.75	0.55	2.83	0.57	2.89	0.58	2.94	0.60
Years of Experience	6.43	3.43	7.92	3.86	9.36	4.19	10.84	4.71	12.65	5.07
Current Unemployment	7.03	2.21	7.26	2.34	6.55	2.38	5.56	2.39	4.61	2.53
Corrected AFQT	0.57	0.76	0.56	0.76	0.57	0.75	0.53	0.78	0.50	0.80
Number of Siblings	2.63	1.67	2.65	1.69	2.68	1.75	2.71	1.77	2.68	1.74
Mother's Schooling	12.90	2.25	12.89	2.21	12.90	2.24	12.84	2.25	12.83	2.22
Father's Schooling	13.67	3.09	13.68	3.08	13.71	3.09	13.68	3.12	13.61	3.10
Average County Tuition	19.80	8.16	19.69	8.21	19.66	8.14	19.78	8.14	19.62	8.24
Distance to College	3.15	9.78	3.16	10.07	3.30	10.19	3.33	10.20	3.10	9.54
Unemployment in 1979	6.29	1.69	6.28	1.68	6.28	1.70	6.29	1.67	6.27	1.66
Some College	0.45	0.50	0.45	0.50	0.44	0.50	0.45	0.50	0.46	0.50
High School ($S = 0$)	$n = 820$		$n = 789$		$n = 750$		$n = 717$		$n = 654$	
Log Wage	2.37	0.60	2.37	0.57	2.41	0.51	2.46	0.54	2.47	0.58
Years of Experience	9.03	3.64	10.56	4.05	12.00	4.48	13.50	4.96	15.22	5.32
Current Unemployment	7.19	2.28	7.40	2.46	6.73	2.61	5.78	2.57	4.80	2.64
Corrected AFQT	-0.47	0.90	-0.48	0.89	-0.49	0.89	-0.49	0.89	-0.50	0.89
Number of Siblings	3.25	2.09	3.24	2.08	3.22	2.02	3.25	2.05	3.30	2.08
Mother's Schooling	11.33	2.13	11.30	2.13	11.25	2.15	11.26	2.15	11.21	2.16
Father's Schooling	11.16	2.98	11.13	2.93	11.08	2.94	11.00	2.91	10.94	2.88
Average County Tuition	20.61	8.31	20.67	8.34	20.62	8.34	20.68	8.32	20.91	8.39
Distance to College	4.23	10.30	4.34	10.44	4.41	10.52	4.47	10.63	4.35	10.45
Unemployment in 1979	6.40	1.69	6.40	1.71	6.37	1.69	6.38	1.71	6.43	1.73
High School Dropout	0.19	0.39	0.17	0.38	0.17	0.37	0.16	0.37	0.17	0.37

Note: Entries in this table are means and standard deviations of variables. In addition, n denotes the sample size of each schooling group. The log wages are 5 year averages of non-missing hourly wages. Years of experience are actual work experience from 1979. Current unemployment is the state unemployment in percentage in the current state of residence. Correct AFQT is schooling-adjusted and normalized to have mean zero in the NLSY population. Parental schooling is measured in years of education. Average county tuition is averages of four year public colleges, in hundreds of dollars, at age 19. Distance to college is distance to any college, in miles, at age 14. Unemployment in 1979 is in percentage. Finally, some college and high school dropout are indicator variables that have value one when an individual belongs to corresponding education groups.

Table 2: Average Derivatives for the College Attendance Logit Model

Variable	Year 1992	Year 1994	Year 1996	Year 1998	Year 2000
Corrected AFQT	0.2206	0.2165	0.2229	0.2109	0.2018
	(0.0140)	(0.0140)	(0.0145)	(0.0144)	(0.0144)
Number of Siblings	-0.0148	-0.0143	-0.0159	-0.0129	-0.0151
	(0.0067)	(0.0066)	(0.0066)	(0.0068)	(0.0070)
Mother's Schooling	0.0236	0.0238	0.0208	0.0144	0.0179
	(0.0063)	(0.0064)	(0.0064)	(0.0065)	(0.0068)
Father's Schooling	0.0231	0.0248	0.0263	0.0313	0.0318
	(0.0043)	(0.0043)	(0.0044)	(0.0046)	(0.0047)
Average County Tuition	-0.0028	-0.0032	-0.0037	-0.0037	-0.0047
	(0.0014)	(0.0014)	(0.0014)	(0.0015)	(0.0015)
Distance to College	-0.0026	-0.0024	-0.0013	-0.0010	-0.0010
	(0.0023)	(0.0021)	(0.0020)	(0.0020)	(0.0024)
Unemployment in 1979	0.0046	0.0067	0.0082	0.0117	0.0070
	(0.0078)	(0.0078)	(0.0079)	(0.0082)	(0.0084)
Test for Instruments	0.113	0.057	0.052	0.060	0.015

Note: For each year, the average derivatives are obtained from a logit regression of college attendance on a constant, linear, quadratic, and interaction terms of variables listed in the table. Standard errors are in parentheses, obtained using asymptotic approximation and the delta method. The last row shows p-values for testing the null hypothesis that the average derivatives of tuition, distance, and unemployment in 1979 are jointly zero.

Table 3: Analysis of Counterfactual Variances of Y_1 and Y_0

Component	Year 1992	Year 1994	Year 1996	Year 1998	Year 2000
Panel A: Variance Decomposition of Y_1					
$E[\text{Var}(U_1|V)]$	0.252	0.240	0.284	0.257	0.297
$\text{Var}[\mu_1(X)]$	0.081	0.058	0.081	0.087	0.059
$\text{Var}[E(U_1|V)]$	0.062	0.042	0.002	0.013	0.008
$E[\text{Var}(U_1|V)]/\text{Var}(Y_1)$	0.638	0.706	0.774	0.721	0.816
Panel B: Variance Decomposition of Y_0					
$E[\text{Var}(U_0|V)]$	0.161	0.172	0.188	0.205	0.229
$\text{Var}[\mu_0(X)]$	0.016	0.019	0.021	0.027	0.030
$\text{Var}[E(U_0|V)]$	0.062	0.010	0.006	0.014	0.044
$E[\text{Var}(U_0|V)]/\text{Var}(Y_0)$	0.674	0.852	0.873	0.833	0.756

Note: The first panel of this table decomposes the variance of Y_1 in a component due to X (second line), another due to V (third line) and a third one due neither to X nor V (first line). The latter represents the variance in Y_1 that is not related with selection and in the fourth line of the panel we report the percentage of the total variance accounted for this component. The second panel presents a similar decomposition for Y_0.

Table 4: Results of Simulating 1980's

Variable	Census 1980	Census 1990	Simulation 1980	Simulation 1990
College Enrollment Rates	0.41	0.55	0.41	0.55
Average Overall Wages	2.60	2.62	2.54	2.57
Average College Wages	2.72	2.76	2.77	2.71
Average High School Wages	2.50	2.43	2.37	2.38
College Premium (OLS)	0.22	0.33	0.40	0.33
90-10 Overall Wages	1.41	1.56	1.28	1.29
90-10 College Wages	1.46	1.57	1.30	1.34
90-10 High School Wages	1.34	1.43	1.13	1.12

Average Counterfactual College Wages at the Margin	2.55
Average Counterfactual High School Wages at the Margin	2.35
Average Treatment Effect at the Margin	0.20

Note: The first two columns present measures of average schooling and characteristics of the wage distribution using Census data from 1980 and 1990 for white males. The second two columns present characteristics of simulated wage distributions from our model under the assumption that the college participation rate is 41% (third column) and 55% (fourth column). The last three lines refer to individuals at the margin, which are defined as those who are induced to go to college by the change of college enrollment rate from 41% to 55%.

Table 5: Appendix Table: Average Derivatives for College and High School Wages

Variable	Year 1992	Year 1994	Year 1996	Year 1998	Year 2000
Panel A: Average Derivatives for College Wages					
Years of Experience	0.008	0.008	0.012	0.009	0.004
	(0.013)	(0.011)	(0.010)	(0.009)	(0.008)
Current Unemployment	-0.000	0.003	0.009	0.004	-0.010
	(0.012)	(0.012)	(0.012)	(0.013)	(0.011)
Corrected AFQT	0.238	0.216	0.195	0.191	0.152
	(0.138)	(0.132)	(0.138)	(0.125)	(0.136)
Number of Siblings	0.015	0.000	-0.017	-0.012	-0.011
	(0.019)	(0.018)	(0.019)	(0.018)	(0.017)
Mother's Schooling	0.023	0.005	0.012	0.008	0.015
	(0.016)	(0.018)	(0.023)	(0.021)	(0.022)
Father's Schooling	0.004	0.008	0.023	0.029	0.020
Panel B: Average Derivatives for High School Wages					
Years of Experience	0.002	0.003	-0.002	-0.001	0.002
	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Current Unemployment	0.004	0.010	0.006	0.013	0.004
	(0.013)	(0.011)	(0.009)	(0.014)	(0.016)
Corrected AFQT	0.058	0.060	-0.021	0.027	-0.049
	(0.110)	(0.104)	(0.091)	(0.084)	(0.102)
Number of Siblings	0.004	-0.001	-0.008	-0.001	-0.004
	(0.015)	(0.015)	(0.016)	(0.022)	(0.022)
Mother's Schooling	0.006	0.007	0.019	0.029	0.021
	(0.016)	(0.017)	(0.017)	(0.020)	(0.023)
Father's Schooling	0.003	0.003	-0.009	-0.018	-0.013
	(0.013)	(0.014)	(0.019)	(0.030)	(0.034)

Note: The average derivatives are obtained from partially linear regressions of log wages using observations with $S = 1$ and $S = 0$, respectively. The X vector in the wage equations are described in section 3.1. The nonparametric component of the partially linear regression is the predicted probability of attending college. The predicted probability is estimated by a logit regression reported in Table 2. Standard errors are in parentheses, obtained from 200 bootstrap replications.

Table 6: Appendix Table: P-values of Tests for Selection

Test	Year 1992	Year 1994	Year 1996	Year 1998	Year 2000
College Wages					
No selection on unobservables	0.011	0.032	0.086	0.021	0.094
No selection on family background and unobservables	0.040	0.074	0.019	0.000	0.002
No selection on AFQT and unobservables	0.000	0.000	0.002	0.000	0.005
No selection on AFQT, family backgound and unobservables	0.000	0.000	0.000	0.000	0.000
LR Test for no selection on unobservables	0.116	0.052	0.033	0.022	0.033
High School Wages					
No selection on unobservables	0.359	0.185	0.206	0.498	0.783
No selection on family background and unobservables	0.225	0.129	0.137	0.270	0.346
No selection on AFQT and unobservables	0.013	0.001	0.003	0.004	0.004
No selection on AFQT, family backgound and unobservables	0.003	0.001	0.001	0.009	0.044
LR Test for no selection on unobservables	0.862	0.000	0.556	0.741	0.343

Note: Entries in the table are p-values of tests for selection for college and high school wages. The first four tests are Hausman-type bootstrap tests, which are based on the mean square difference between fitted values of the restricted and unrestricted models. For tests for no selection on unobservables, the restricted models are linear regression models with specifications that are identical to the parametric parts of partially linear models reported in table 5. For all tests, the unrestricted models are linear regression models with the same specifications above along with the following additional regressors: linear, quadratic, and interaction terms of instruments (tuition, distance, unemployment in 1979) and interactions between instruments and AFQT and family background variables (number of siblings, mother's education, and father's education). For tests for no selection on some observables and unobservables, the restricted models are linear regression models with similar specifications except that corresponding observables are omitted from the specifications. Each p-value is obtained (without Studentization) by wild bootstrap with 999 bootstrap replications. Finally, for comparison, the likelihood ratio (LR) test is carried out via the maximum likelihood estimation using the normal selection model.
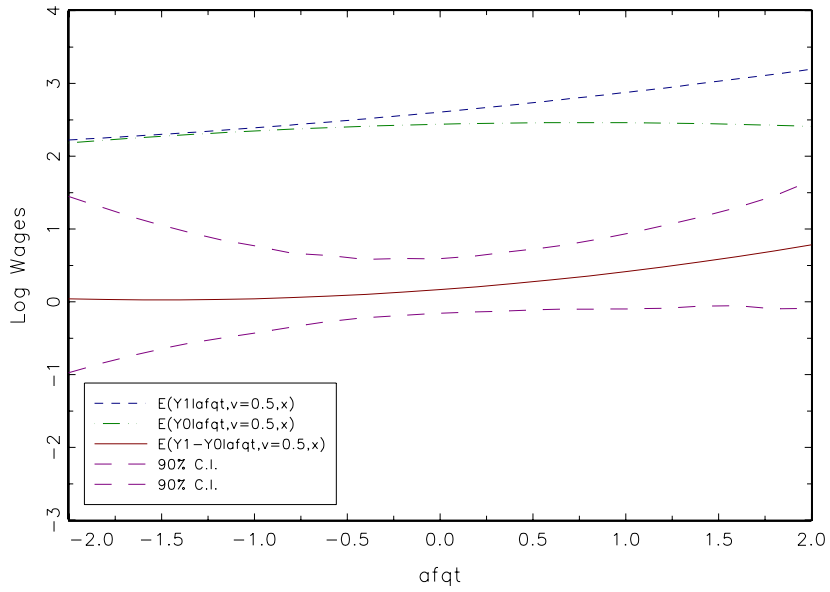
44
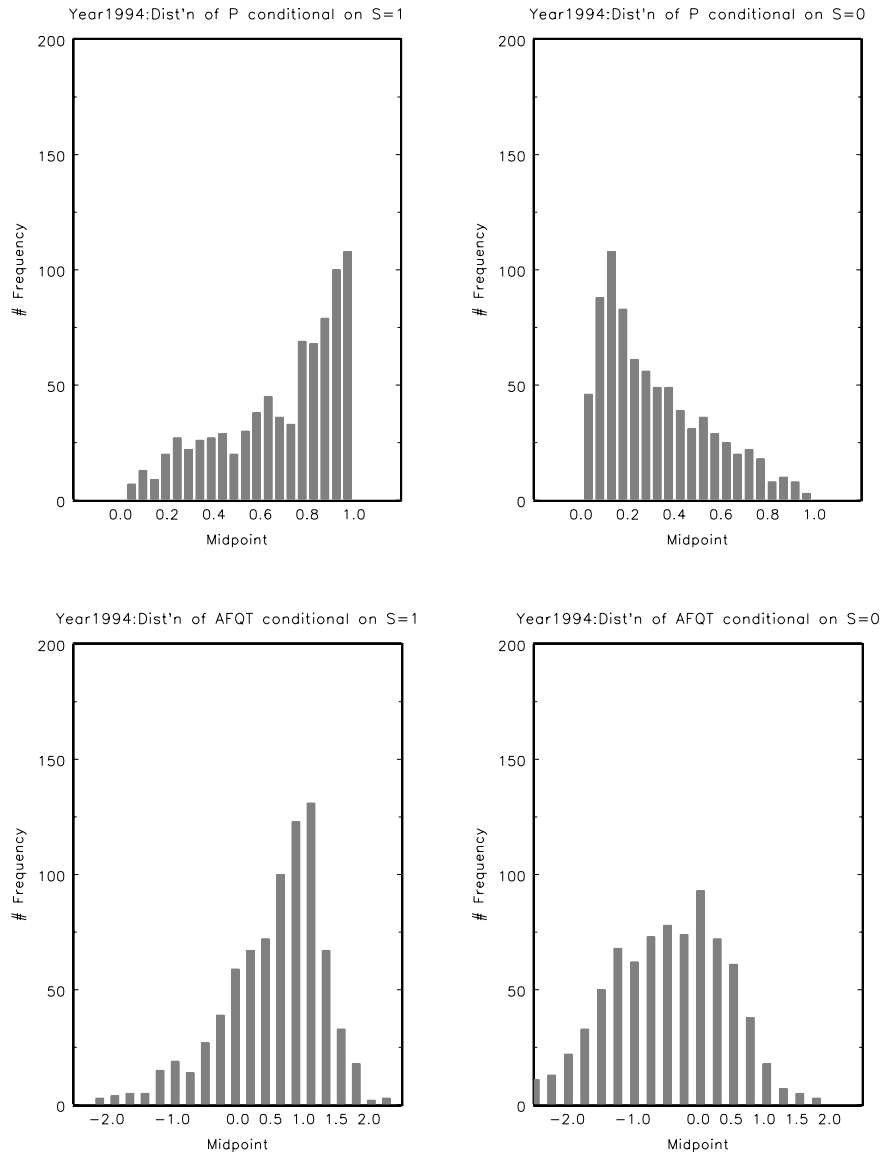
Figure 1: Support of P and AFQT (Year 1994)

45

Figure 2: MTE as a Function of AFQT (Year 1994)

Year1994: Decomposition of E(Y1−Y0|afqt,v=0.5,x)

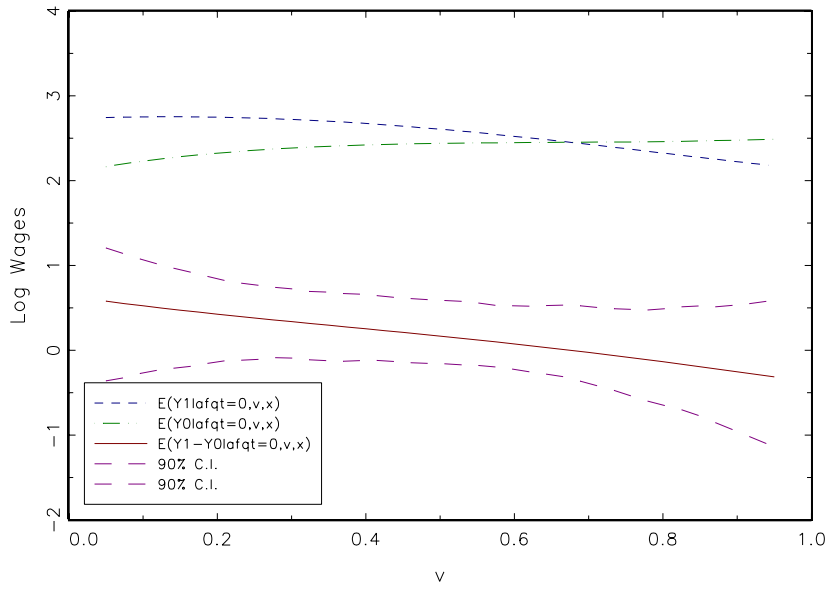Year1994: Decomposition of E(Y1−Y0|afqt=0,v,x)

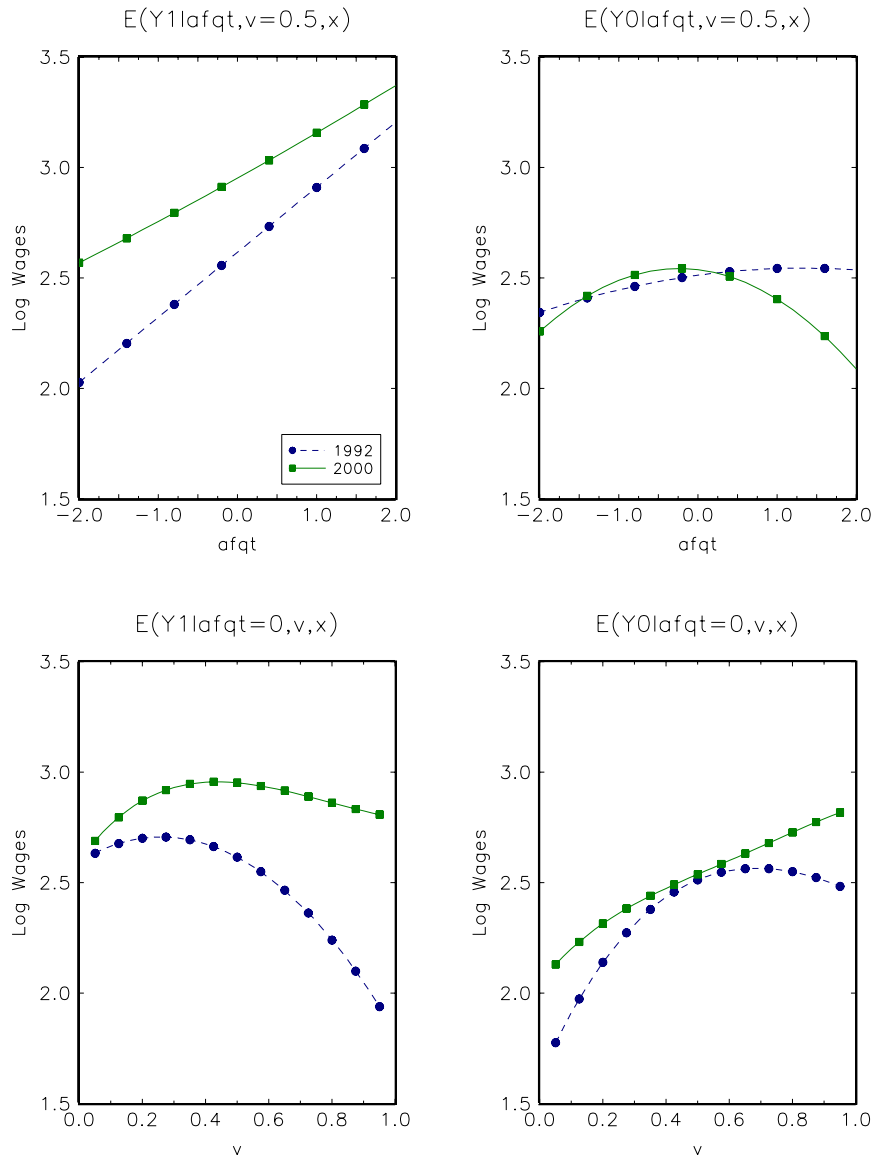Figure 3: MTE as a Function of V (Year 1994)

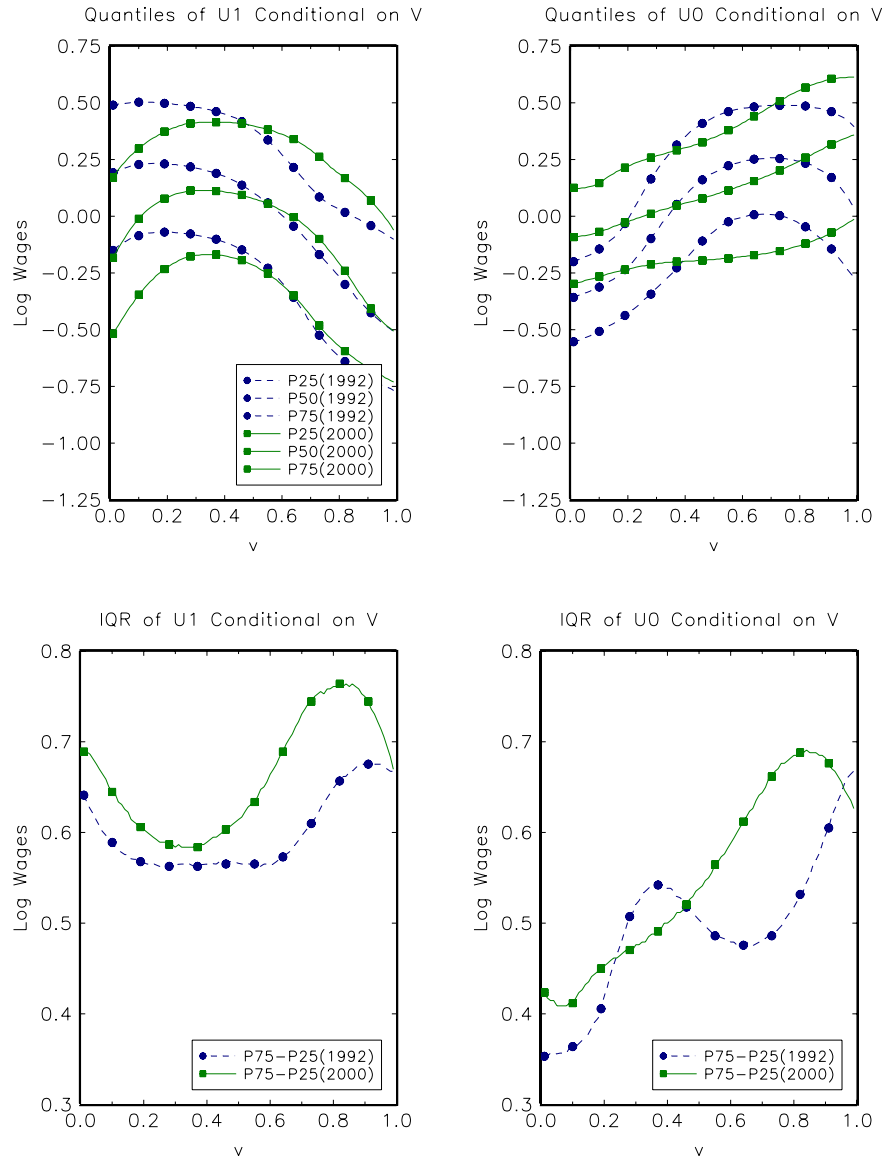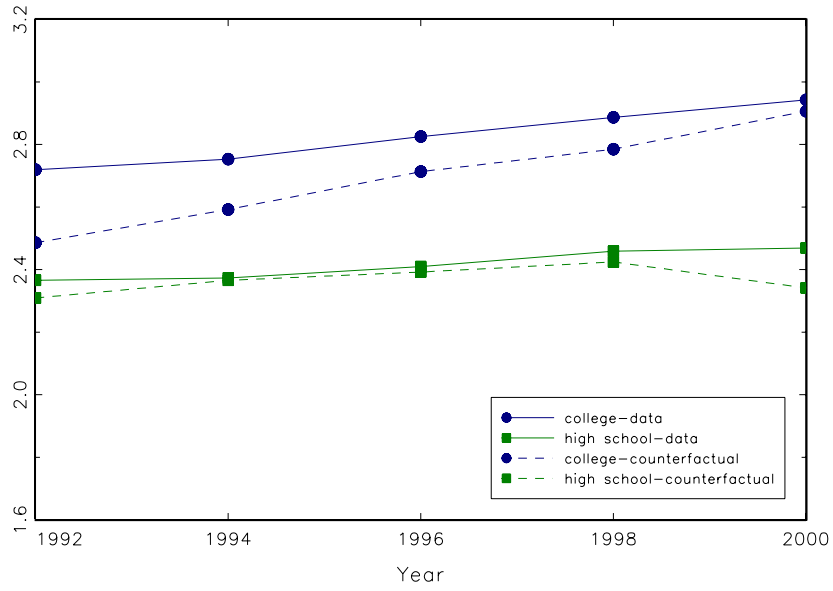Figure 4: $E[Y_1|X,V]$ and $E[Y_0|X,V]$ (Years 1992 and 2000)

Figure 5: $Q[U_1|V]$ and $Q[U_0|V]$ (Years 1992 and 2000)

48

Figure 6: Counterfactual Means of College and High School Wages

Counterfactual and Conditional Means of College and High School

Return to College: ATE, TT, TUT, and OLS
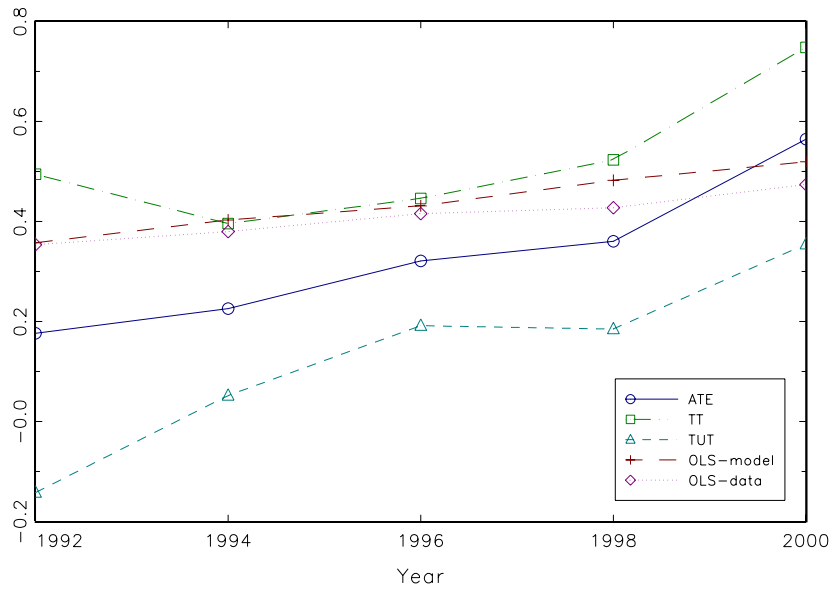
Figure 7: Treatment Effect Parameters

49
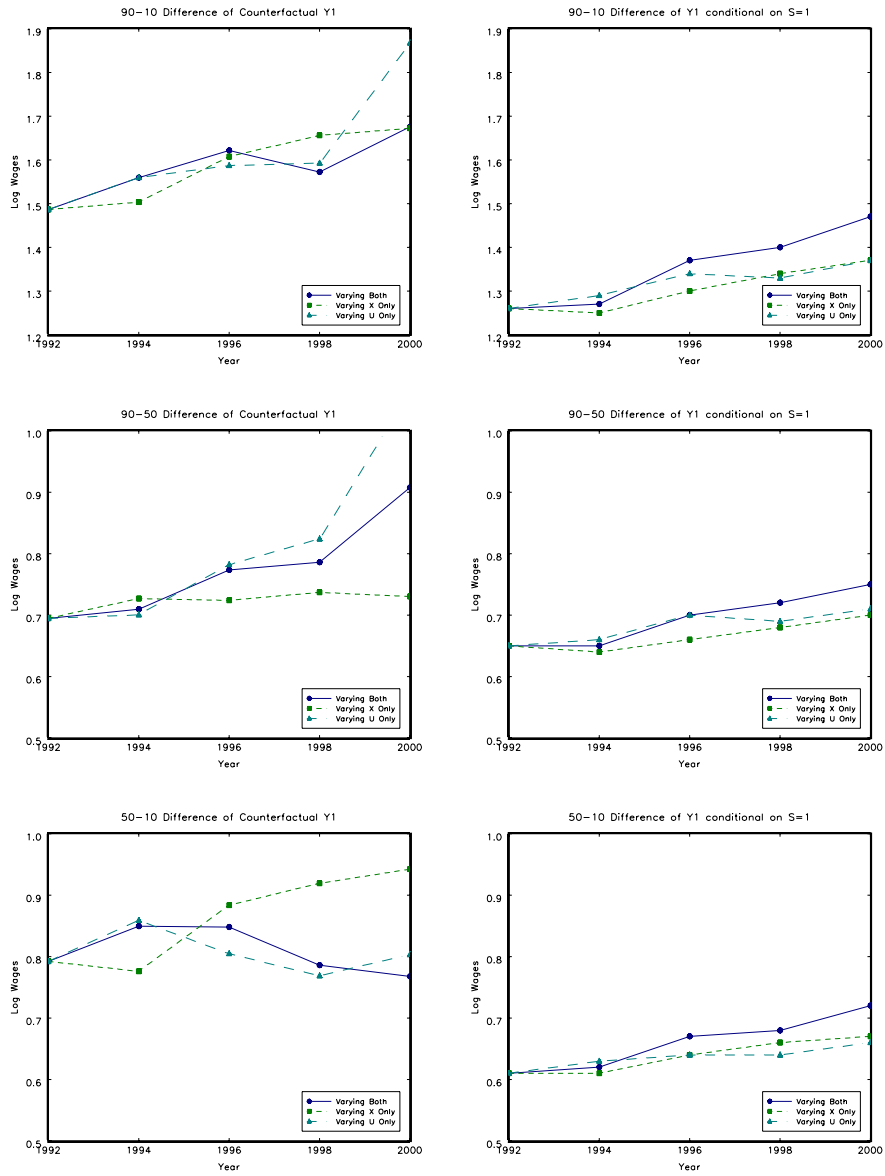
Figure 8: Counterfactual Changes in Within Group Inequality (College)
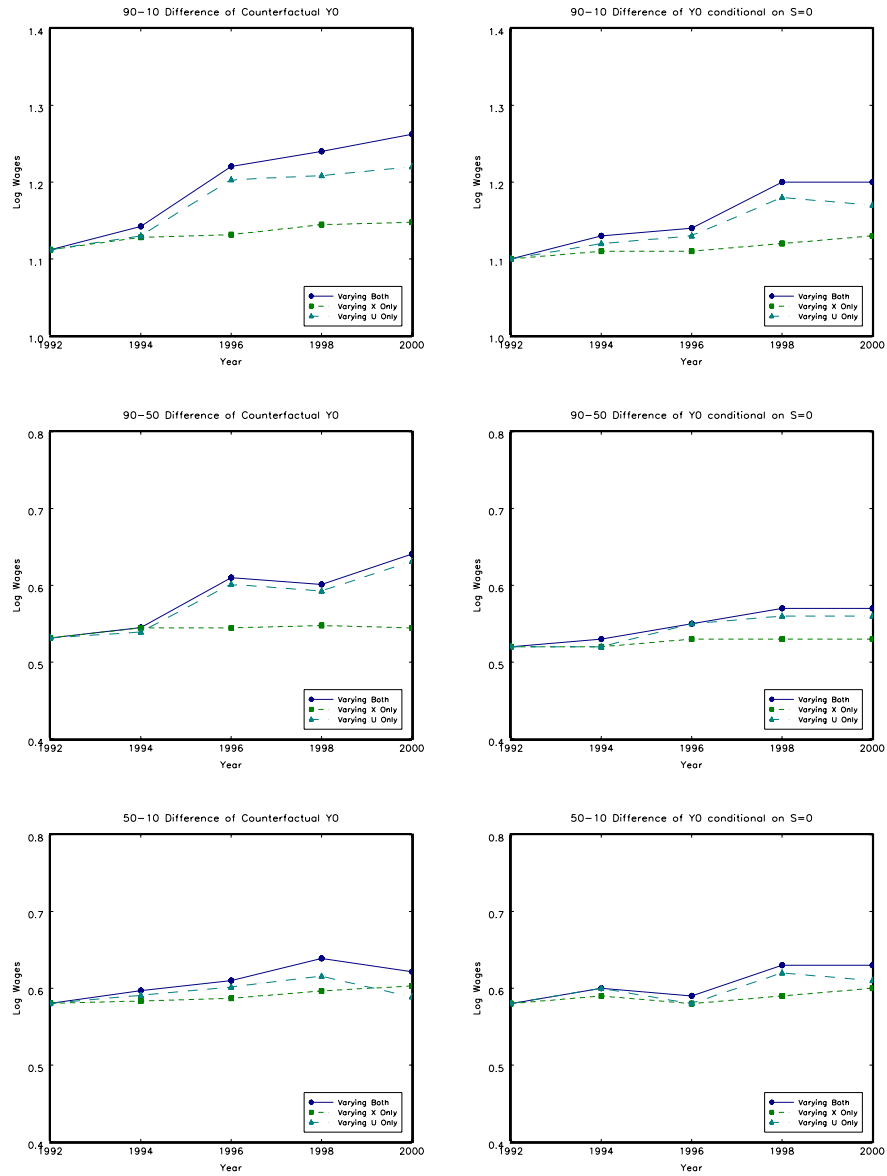
Figure 9: Counterfactual Changes in Within Group Inequality (High School)
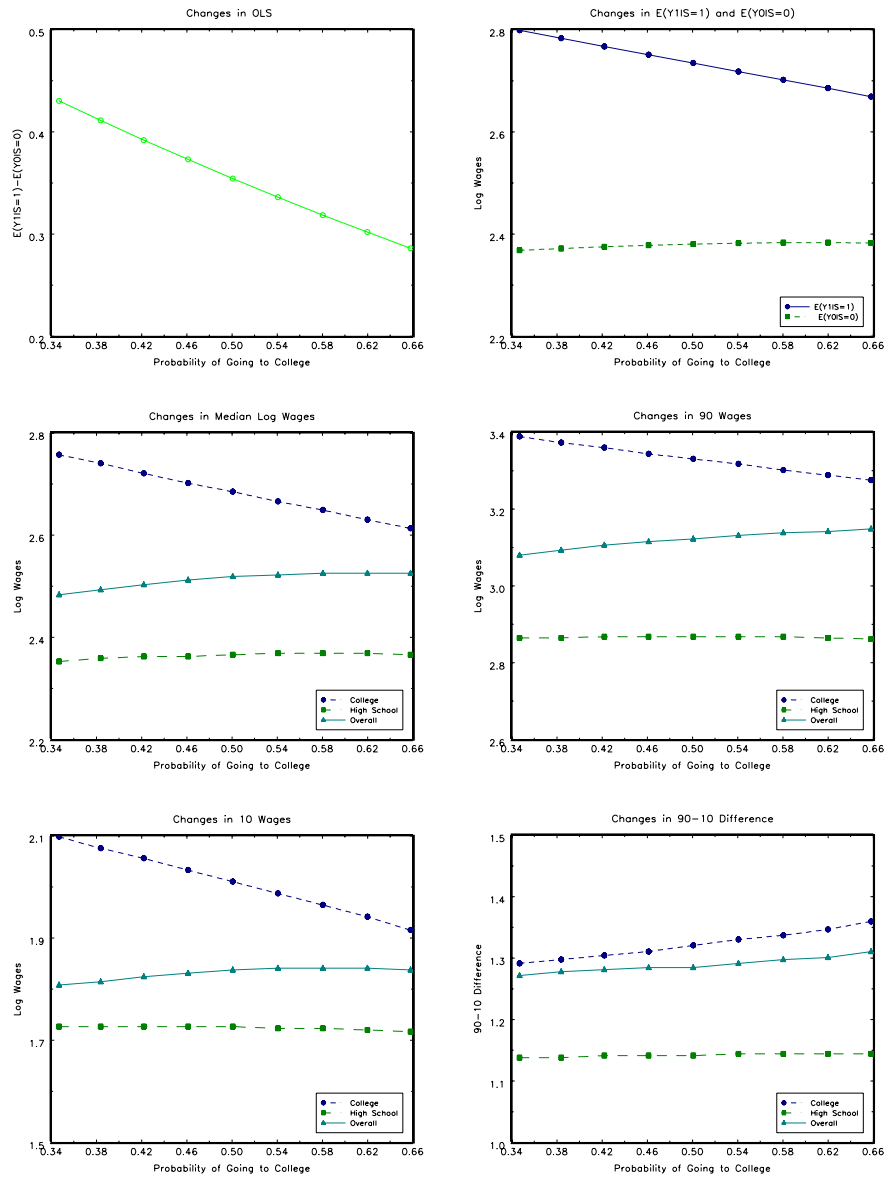
Figure 10: Simulation Results of Wage Distributions of Different Compositions of Education Groups

Log Hourly Wage Distributions for the NLSY and the CPS
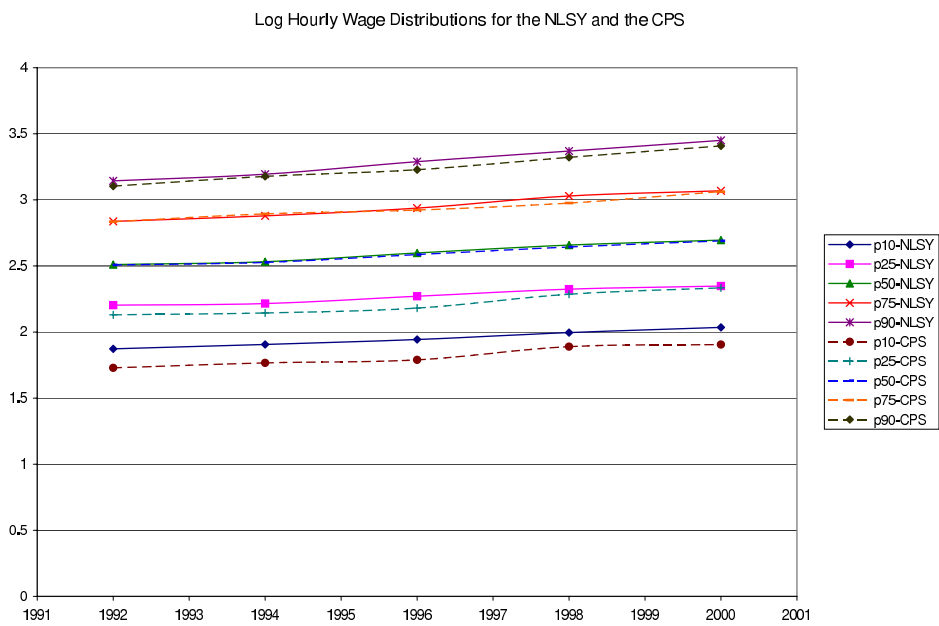
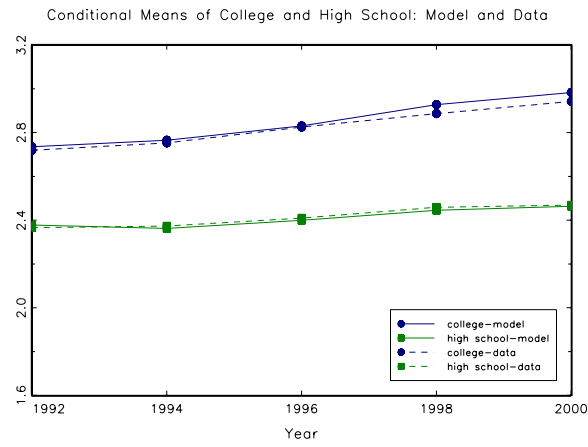Figure 11: Appendix Figure: Trends of Inequality in CPS and NLSY in the 1990s for the Same Cohorts
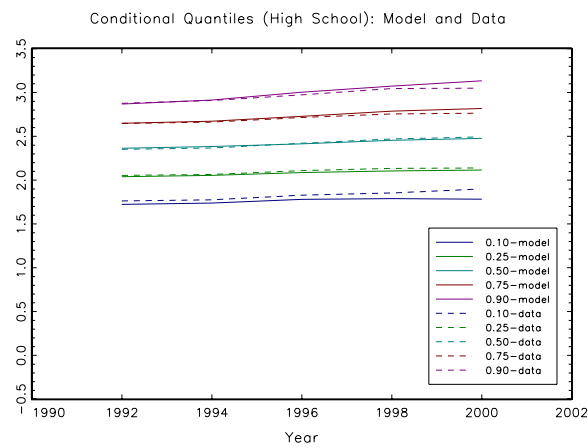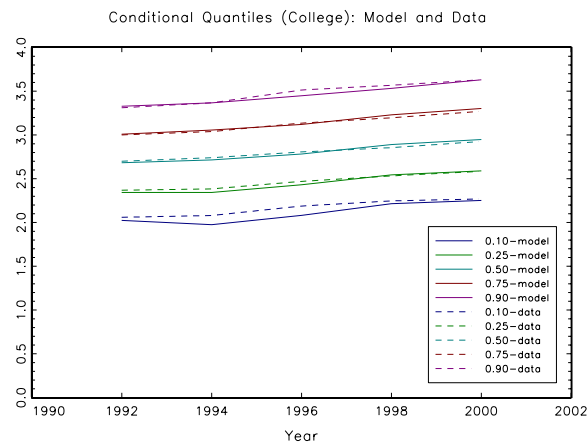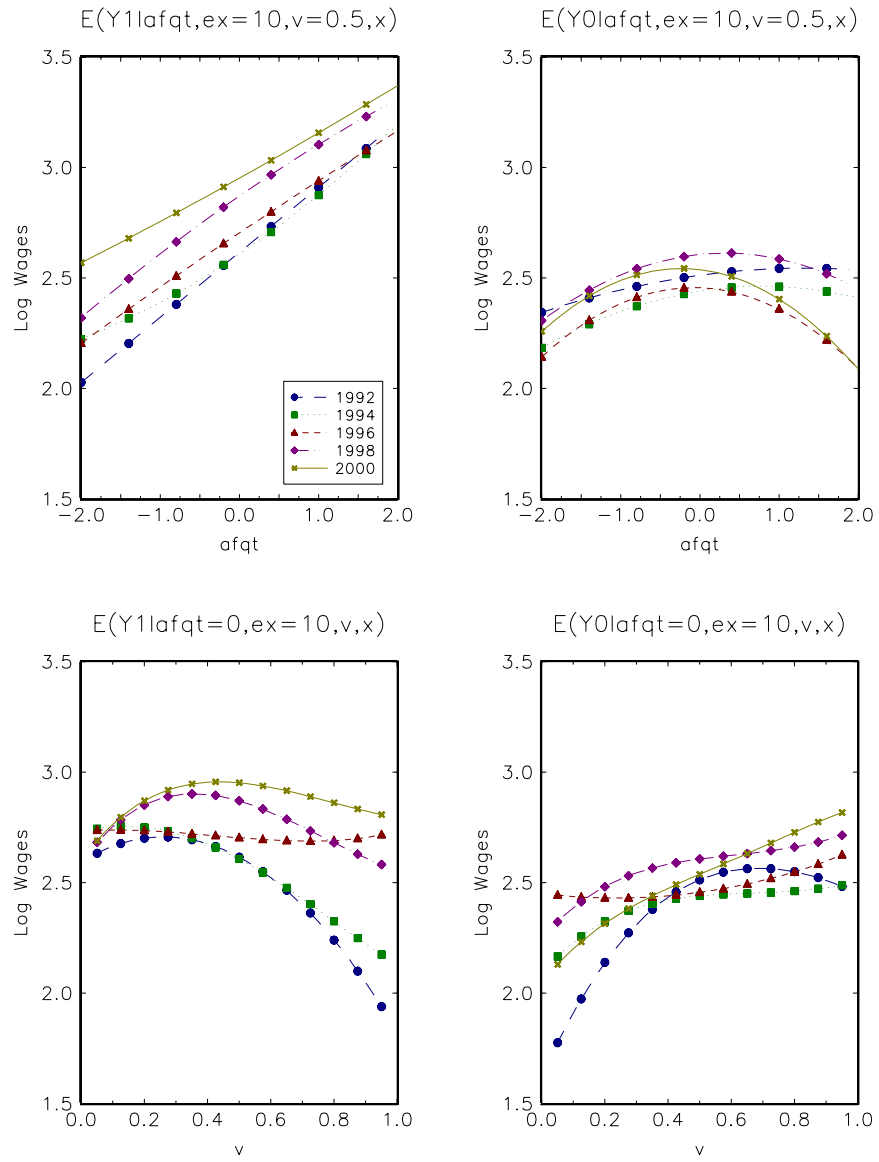
Figure 12: Appendix Figure: A Goodness of Fit Check

Figure 13: Appendix Figure: $E[Y_1|X, V]$ and $E[Y_0|X, V]$ (All Years)