

Ability, Sorting and Wage Inequality

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

May 5, 2005

Incomplete and Preliminary Draft

1 Introduction

Wage inequality has increased dramatically over the last 30 to 40 years in the US and in the UK (see, for example, Autor and Katz, 1999). Economists have devoted a substantial amount of work to the documentation and interpretation of the evolution of wage inequality and several of its dimensions, such as the skill premium, or racial and gender wage gaps. One variable whose role is often stressed is schooling. Schooling is an important determinant of wages and therefore it is likely to substantially affect inequality.

In the large empirical literature on the effect of schooling on wages, usually labeled the return to schooling, it is emphasized that this parameter varies across individuals and that individuals sort themselves non-randomly across schooling levels (see, for example, Card, 1999, and Carneiro, Heckman and Vytlačil, 2003). As a result, considerable care is needed both in the estimation of such returns and in the interpretation of standard estimates. In contrast, in standard studies of inequality, such as Katz and Murphy (1992) or Card and Lemieux (2001), the role of heterogeneity in returns and self-selection into schooling is often ignored.

On the other hand, the role of heterogeneity and self selection into employment has been seriously considered in the study of inequality. For example, Heckman and Todd (2003) show that accounting for differential labor force participation rates of blacks and whites dramatically reduces black-white wage convergence in the last 40 years, because of selective dropout of blacks from the labor force.

*We thank participants in seminars at the 2005 AEA meetings in Philadelphia, Norwegian School of Economics, Kent, Maynooth and Ente Einaudi for useful comments.

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. Blundell, Reed and Stoker (2003) study trends in wage growth and in the returns to schooling in the UK and conclude that, accounting for selective dropout from the labor force, most of the aggregate wage growth in the UK over the last 20 years can be accounted for an increase in the unemployment of individuals in the bottom of the wage distribution. 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. In summary, selection into employment has been shown to be clearly important in the study of the determinants of inequality. However, the treatment of selection into schooling in this literature has not been as rigorous, generating a striking asymmetry in the way we account for selection into employment and into schooling in standard studies of the evolution of wages and wage inequality.

There exist, however, a few exceptions. Papers such as Heckman, Lochner and Taber (1998), Chay and Lee (2000), Taber (2001) and Juhn, Kim and Vella (2005) do account for heterogeneity and selection in the study of different dimensions of inequality. In particular, the latter paper aims to study the role of changes in the composition of schooling groups on least squares measures of the returns to schooling. The authors examine whether an increase in the college population leads to a decrease in the average quality of the college graduate. They find significant composition effects on the evolution of returns, but suggest that these are unlikely to explain recent changes in standard measures of the returns to schooling, although they are probably important in explaining medium to long-run movements in this parameter. In a study of the evolution of inequality in the UK, Gosling, Machin and Meghir (2000) argue that cohort effects drive most of its recent increase. Among other things, they suggest that differential composition of education groups across cohorts may explain partly such cohort effects, but they do not investigate the issue directly. In another recent paper, Acemoglu (2002) also points out that composition effects are potentially important to explain movements in inequality, but that evidence from Juhn, Murphy and Pierce (1993) suggests that such effects are unlikely to play a big role. Juhn, Murphy and Pierce (1993) show that across cohort changes in 90-10 wage differentials are very similar to across age changes in 90-10 wage differentials, suggesting that changes in the composition of ability in the population are not likely

to be important.

Our paper contributes to this literature by analyzing the relationship between schooling and inequality accounting for heterogeneity and self-selection into schooling. We estimate a semi-parametric selection model to study the evolution of inequality in the 1990s for a group of white males surveyed by the National Longitudinal Survey of Youth of 1979 (NLSY). Even though the sample consists of (virtually) a single cohort of individuals, its study allows us to draw broader lessons for the study of inequality and its changes over time. We show that accounting for selection has critical importance in a variety of different dimensions.

Furthermore, we complement our study with an analysis of Census data which also suggests that the issues addressed in this paper are relevant for the interpretation of the trends in inequality. In particular, we revisit Card and Lemieux's (2001) argument that the recent increase in the returns to schooling for young individuals is due to a slowdown in the growth of educational attainment for younger cohorts, causing supply not to keep up with a rising demand for skill, leading to an increase in skill prices. In doing so we argue that the way composition effects are analyzed explicitly in Juhn, Kim and Vella (2005) and implicitly in Juhn, Murphy and Pierce (1993) does not provide us with a complete picture of the influence of sorting on the evolution wage inequality. First, Juhn, Kim and Vella (2005) look for composition effects on an OLS measure of the return to schooling which, as is well known, does not measure the price of schooling for anyone in the economy since it is estimated by comparing non-comparable people. Moreover, as it is made clear in this paper, *a priori* it is not clear how changes in the composition of schooling across cohort affect the OLS estimate of the return to schooling. Second, Juhn, Murphy and Pierce (1993) look at the evolution of 90-10 differentials in the overall wage distribution (before and after residualizing it on several observable variables) while we emphasize that, to analyze composition effects, one should examine the evolution of the wage distribution of each skill group separately.

Two main concerns motivate us to study this problem. First, changes in inequality are usually analyzed and interpreted in a standard demand and supply framework (see, for example, Katz and Murphy, 1992, or Card and Lemieux, 2001). For example, the recent increase in measured returns to schooling (the standard measure of between group inequality) is attributed to an increase in skill prices, which is due to a combination of rising demand for skill and slowdown in the growth of educational attainment in the population (Katz and Murphy, 1992, Card and Lemieux, 2001). But

can we really interpret this data as evidence for changes in skill prices in the economy?

Take the following simple example (the model we develop in this paper is much more general). Suppose there are only two schooling groups, high school and college ($S = 1$ and $S = 0$, respectively). Let:

$$\begin{aligned}\ln Y_{1it} &= \alpha_{1t} + \gamma_{1t}A_i \\ \ln Y_{0it} &= \alpha_{0t} + \gamma_{0t}A_i\end{aligned}$$

where $\ln Y_{1it}$ is the logarithm of the wage of individual i at time t if he goes to college, and $\ln Y_{0it}$ is the corresponding quantity if he does not go to college. Assume A_i is ability and it is unobserved. The return to college for individual i at time t is

$$\beta_{it} = \alpha_{1t} - \alpha_{0t} + (\gamma_{1t} - \gamma_{0t}) A_i.$$

In order to compute the average return to schooling for a particular group of individuals we compare their counterfactual average wages in each schooling level. For example, the average return to college for individuals who enrolled in college is:

$$E(Y_{1it} - Y_{0it} | S_{it} = 1) = \alpha_{1t} - \alpha_{0t} + (\gamma_{1t} - \gamma_{0t}) E(A_i | S_{it} = 1)$$

The OLS (least squares) estimate of the average return which is used in most analyses of inequality is

$$\begin{aligned}\beta_t^{OLS} &= E(Y_{1it} | S_{it} = 1) - E(Y_{0it} | S_{it} = 0) \\ &= \alpha_{1t} - \alpha_{0t} + \gamma_{1t}E(A_i | S_{it} = 1) - \gamma_{0t}E(A_i | S_{it} = 0).\end{aligned}$$

This expression illustrates what is emphasized in the returns to schooling literature: β_t^{OLS} does not measure the return to schooling for any individual in the economy since it compares wages in college and in high school of individuals who are not comparable: this return is obtained by comparing individuals in the $S = 1$ group with individuals in the $S = 0$ group. In an extreme case, even if β_{it} does not change over time for any individual in the economy, β_t^{OLS} can change over time for two

main reasons:

a) If the schooling attainment of the population changes over time then $E(A|S_t = 1)$ and $E(A|S_t = 0)$ can change over time. Such *composition changes* lead to changes in β_t^{OLS} .

b) Even if $E(A|S_t = 1)$ and $E(A|S_t = 0)$ do not change over time, $\gamma_{1t}E(A|S_t = 1) - \gamma_{0t}E(A|S_t = 0)$, which is usually called *selection bias*, can also change over time if, for example, γ_{1t} and γ_{0t} are not constant over time, even if their difference ($\gamma_{1t} - \gamma_{0t}$) is constant over time.

More generally, the change in the OLS measure of the return to schooling may be a very misleading indicator of changes in the prices of skill in the economy. Even though this example focused on a measure of between group inequality, a similar reasoning applies to the study of the evolution of within group wage inequality.

The second concern that motivates the importance of our work is the following. Using the standard setup it is not possible to understand the effect of education policy on wage inequality. In a partial equilibrium framework where skill prices are fixed, if we want to know the impact on inequality of a policy that increases schooling attainment we need to know what is the return to schooling for those individuals induced to increase their schooling by the policy, how their exit from the $S = 0$ group changes $f(Y_0|S = 0)$ and how their entry into the $S = 1$ group changes $f(Y_1|S = 1)$, where $f(\cdot|S = s)$ is the density of wages for individuals with schooling equal to $S = s$, where $s = 0, 1$.¹ Ferreira and Leite (2001) apply these ideas to the simulation of the effects of an education reform in the Brazilian state of Ceara. The framework developed in this paper is considerably more flexible than theirs and the model is estimated on better data.²

We illustrate the importance of both these issues in an analysis of inequality in the 1990s for white males in the NLSY. In order to do this we estimate a semiparametric selection model using the method of local instrumental variables (see Heckman and Vytlacil, 2000, 2004). We extend this method to the estimation of marginal distributions of outcomes in each sector. Our work is closely to the literature on the estimation of quantile treatment effects using instrumental variables (see, in particular, Abadie, Angrist and Imbens, 2003, Imbens and Rubin, 1996). This paper also builds heavily on our previous work (Carneiro and Lee, 2004). Carneiro, Hansen and Heckman (2003) and Cunha, Heckman and Navarro (2004a,b) also estimate distributions of income and wages allowing

¹In a general equilibrium framework, as in Heckman, Lochner and Taber (2000), we also need to know how skill prices change when there are changes in the level of schooling of the population.

²Ferreira and Leite (2001) allow for multiple years of schooling while we allow for only two levels of schooling in our current setup.

for multiple schooling specific abilities and self-selection into schooling. The methods they use are different than the ones we employ, but the underlying assumptions of their model are very similar to ours. However, their empirical analyses are not applied to changes in inequality over time.

In the next section we present the model which underlies our empirical work. With this framework in the background, before presenting our work in the cohort of individuals surveyed in the NLSY, we briefly revisit Card and Lemieux's (2001) analysis of the evolution of the return to schooling using Census data for 1980 and 2000. In section 4 we apply our model to the study of the evolution of wage inequality in the 1990s for white males in the NLSY. In section 5 we examine the role of changes in educational attainment (due, for example, to education policy) on wage inequality. In section 6 we conclude our paper.

2 Model

The econometric model used in this paper is based on Heckman and Vytlacil (2004), Carneiro, Heckman and Vytlacil (2003) and Carneiro and Lee (2004). Our point of departure is the binary treatment model that is standard in the programme evaluation literature. Suppose there is a set of individuals which we follow over time (in our case, the NLSY respondents, which correspond to a particular set of cohorts of the US population). Let Y_{1t} and Y_{0t} be their potential outcomes in two states, 1 and 0, at time t . In this paper Y_{1t} and Y_{0t} are the potential log wages in college and high school for several years in the 1990s. We assume

$$\begin{aligned} Y_{1t} &= \mu_{1t}(X_t, \beta_{1t}) + U_{1t}, \quad \text{and} \\ Y_{0t} &= \mu_{0t}(X_t, \beta_{0t}) + U_{0t}, \end{aligned} \tag{1}$$

where X_t is a random vector influencing potential outcomes, the functional forms of μ_{1t} and μ_{0t} are specified up to unknown finite dimensional parameters β_{1t} and β_{0t} , respectively, and U_{1t} and U_{0t} are unobserved random variables for each t . We assume that individuals choose to be in state 1 or 0 (prior to the realizations of outcomes) according to the following equation:

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

where Z is a random vector influencing the decision equation, μ_S is a parametric 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.³ For each individual at time t , the observed outcome Y_t is

$$Y_t = SY_{1t} + (1 - S)Y_{0t}.$$

The set of variables in X_t can be a subset of Z . In this paper we assume that there is at least one variable in Z that is not in X_t (exclusion restriction). We also assume that (U_{1t}, U_S) and (U_{0t}, U_S) are independent of (Z, X_t) .⁴ We can rewrite (2) as:

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

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

In this paper we estimate functionals of the marginal distributions of Y_{1t} and Y_{0t} , in particular the means and quantiles of Y_{1t} and Y_{0t} , conditional on X_t and V_t . As by-products, we also estimate average treatment effect (ATE), treatment effect on the treated (TT), and treatment effect on the untreated (TUT). Estimation of counterfactual distributions and treatment effect parameters allows us to gain important insights into the role of heterogeneity and selection into schooling for the changes in wage distribution.

The following lemma, which is proved in Carneiro and Lee (2004), gives identification of the model that will be estimated in this paper.

Lemma 1 *Consider a semiparametric selection model given by (1) and (2). Let $V = F_{U_S}[U_S]$ and $P = F_{U_S}[\mu_S(Z)]$. Assume that (1) $\mu_S(Z)$ is a nondegenerate random variable conditional on X_t ; (2) (U_{1t}, U_S) and (U_{0t}, U_S) are independent of (Z, X_t) ; (3) The distributions of U_S and $\mu_S(Z)$ are absolutely continuous with respect to Lebesgue measure; (4) $E|Y_{1t}| < \infty$ and $E|Y_{0t}| < \infty$; (5)*

³Carneiro, Heckman and Vytlačil (2003) use this model to examine heterogeneity in the returns to college and present an economic model that can justify the specification in (2).

⁴This assumption is useful to reduce the dimension of the estimation problem, thereby helping to achieve good precision of estimates. See Carneiro and Lee (2004) for results on the identification of a nonparametric, nonseparable version of the model (1) under weaker assumptions than those imposed in this paper.

$0 < Pr(S = 1|X_t) < 1$; and (6) $E[U_{1t}|P = p, S = 1]$, $E[U_{0t}|P = p, S = 0]$, $Pr[U_{1t}|P = p, S = 1]$ and $Pr[U_{0t}|P = p, S = 0]$ are continuously differentiable with respect to p . Then the expectations, quantiles, and marginal distributions of Y_1 and Y_0 conditional on $X = x$ and $V = v$ are identified.

This Lemma is a simple extension of the local instrumental variables framework of Heckman and Vytlacil (2004) and it is useful to study wage inequality taking into account of heterogeneity and self-selection into schooling.

2.1 Interpreting Changes in Inequality and Consequences of Selection into Schooling for the Study of Wage Inequality

We study the evolution of wage inequality using the model in the previous section. In particular, we analyze between group inequality and within group inequality accounting for heterogeneity and self-selection into schooling.

First, consider between group inequality. A commonly used explanation for changes in between group inequality is changes in relative prices of different skills. Therefore, to analyze between group inequality properly, it is important to estimate relative skill prices and their evolution over time allowing for heterogeneity and self-selection into schooling. In the literature on inequality, the OLS estimate of the return to education is often used as a measure of between group inequality and interpreted as the relative price of skilled labor (see, for example, Katz and Murphy, 1992, or Card and Lemieux, 1998). However, this parameter is not a “return” for any individual in the economy. The empirical literature on the returns to schooling makes this very clear. Even if the return to schooling does not vary in the population, the OLS estimate is contaminated by selection bias. If the return to schooling varies in the population, as in our paper, then there is a return to schooling for each individual in the economy and a variety of average returns can be defined (see, for example, Carneiro, Heckman and Vytlacil, 2003). In this respect, we estimate different measures of average returns and study their changes in 1990s.

Second, consider within group inequality. As in between group inequality, changes in commonly used measures (e.g., 90-10 wage differentials of observed college and high school wages) can be highly misleading if one is interested in changes in the distribution of each skill sector. Changes in the observed distribution of each education sector can be affected both by changes in the composition

of education groups and by changes in the distribution of underlying skills.

In short, accounting for selection into schooling is essential if we want to understand the impact of education on wage inequality. Decompositions of the Oaxaca-Blinder type and possible extensions (Juhn, Murphy and Pierce, 1993; DiNardo, Fortin and Lemieux, 1996) are usually used to assess the impact of different variables (gender, race, age, education, union, etc) on inequality. It is often assumed that all the variables in the conditioning set are exogenous. Assuming that education is exogenous is at odds with most of the literature on the returns to schooling (Card, 1999, 2000). Once we estimate the model specified in the previous section we can ask questions such as: what would be the overall level of inequality if every individual attended college? what would be the overall level of inequality if no individual attended college? what would be the overall level of inequality if we designed a policy that increased college enrolment by 10%? Answering such questions allows us to describe the potential role of education in poverty.

2.2 Treatment Effect Parameters and Counterfactual Means of Wage Distributions

This section gives a description of estimators of $E[Y_{1t}]$, $E[Y_{0t}]$, $E[Y_{1t}|S = 1]$, $E[Y_{1t}|S = 0]$, $E[Y_{0t}|S = 1]$, and $E[Y_{0t}|S = 0]$. Our estimation method is based on estimators of $E[Y_{1t}|X_t = x_t, V = v]$ and $E[Y_{0t}|X_t = x_t, V = v]$. As suggested by Heckman and Vytlacil (2000, 2004), we can estimate $E[Y_{1t}]$, $E[Y_{1t}|S = 1]$, $E[Y_{1t}|S = 0]$, $E[Y_{0t}|S = 1]$, and $E[Y_{0t}|S = 0]$ by integrating estimators of $E[Y_{1t}|X_t = x_t, V = v]$ and $E[Y_{0t}|X_t = x_t, V = v]$ out with some suitable weights, which will be given below. Using estimates of these conditional expectations, the treatment effect parameters can be estimated by the following simple formulae:

$$\text{ATE}(t) = E[Y_{1t} - Y_{0t}] = E[Y_{1t}] - E[Y_{0t}],$$

$$\text{TT}(t) = E[Y_{1t} - Y_{0t}|S = 1] = E[Y_{1t}|S = 1] - E[Y_{0t}|S = 1],$$

$$\text{TUT}(t) = E[Y_{1t} - Y_{0t}|S = 0] = E[Y_{1t}|S = 0] - E[Y_{0t}|S = 0],$$

and

$$\text{OLS}(t) = E[Y_{1t}|S = 1] - E[Y_{0t}|S = 0]$$

for each t . Note that $E[Y_{1t}|S = 1]$ and $E[Y_{0t}|S = 0]$ 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 $E[Y_{1t}|S = 1]$ and $E[Y_{0t}|S = 0]$ provides some goodness-of-fit check of our model. Also notice that $E[Y_{1t}|S = 0]$ and $E[Y_{0t}|S = 1]$ are counterfactual expectations and it is in general impossible to estimate them using data only.⁵

We start by describing estimators of $E[Y_{1t}|X_t = x_t, V = v]$ and $E[Y_{0t}|X_t = x_t, V = v]$. Under the assumption that (U_{1t}, U_S) and (U_{0t}, U_S) are independent of (Z, X_t) ,

$$E[Y_{1t}|X_t = x_t, V = v] = \mu_{1t}(x_t, \beta_{1t}) + E[U_{1t}|V = v],$$

and

$$E[Y_{0t}|X_t = x_t, V = v] = \mu_{0t}(x_t, \beta_{0t}) + E[U_{0t}|V = v],$$

Thus, estimates of $E[Y_{1t}|X_t = x_t, V = v]$ and $E[Y_{0t}|X_t = x_t, V = v]$ can be obtained by estimating β_{1t} , β_{0t} , $E[U_{1t}|V = v]$, and $E[U_{0t}|V = v]$.

Following Carneiro and Lee (2004), β_{1t} and β_{0t} are estimated by applying a semiparametric version of the sample selection estimator of Das, Newey, and Vella (2003) to each cross section.⁶ To obtain nonparametric estimators of $E[U_{1t}|V = v]$ and $E[U_{0t}|V = v]$, we apply an identification result of Lemma 1 of Carneiro and Lee (2004) to our model (1).⁷ Specifically, it can be shown that under the assumptions give in Lemma 1,

$$E[U_{1t}|V = v] = E[U_{1t}|P = v, S = 1] + v \frac{\partial E[U_{1t}|P = v, S = 1]}{\partial p} \quad \text{and} \quad (3)$$

$$E[U_{0t}|V = v] = E[U_{0t}|P = v, S = 0] - (1 - v) \frac{\partial E[U_{0t}|P = v, S = 0]}{\partial p}. \quad (4)$$

This suggests that $E[U_{1t}|V = v]$ and $E[U_{0t}|V = v]$ can be estimated by sample analogs of the right-hand sides of equations (3) and (4), respectively. Local polynomial estimation is used in this paper to estimate $E(U_{1t}|P = v, S = 1)$, $E(U_{0t}|P = v, S = 0)$ and their partial derivatives with

⁵Of course, under the assumption of ignorable selection, $E[Y_{1t}|S = 1] = E[Y_{1t}|S = 0]$ and $E[Y_{0t}|S = 1] = E[Y_{0t}|S = 0]$, so that we can estimate $E[Y_{1t}|S = 0]$ and $E[Y_{0t}|S = 1]$ by sample analogs of $E[Y_{1t}|S = 1]$ and $E[Y_{0t}|S = 0]$, respectively. However, we do not intend to impose the assumption of ignorable selection *a priori*.

⁶The resulting model is a partially linear model with generated regressors for the nonparametric component. We use a Robinson (1988)-type estimator rather than a series estimator, which is used in Das, Newey, and Vella (2003). See Carneiro and Lee (2004) for details.

⁷Lemma 1 of Carneiro and Lee (2004) is a simple extension of the local instrumental variables framework of Heckman and Vytlacil (2004).

respect to P . See Carneiro and Lee (2004) for a detailed description of the method for estimating $E[Y_{1t}|X_t = x_t, V = v]$ and $[Y_{0t}|X_t = x_t, V = v]$.

Once $E[Y_{1t}|X_t = x_t, V = v]$ and $[Y_{0t}|X_t = x_t, V = v]$ are estimated, we obtain estimators of $E[Y_{1t}]$, $E[Y_{0t}]$, $E[Y_{1t}|S = 1]$, $E[Y_{1t}|S = 0]$, $E[Y_{0t}|S = 1]$, and $E[Y_{0t}|S = 0]$ by the sample analogs of the following formulae:⁸

$$\begin{aligned}
E[Y_{1t}] &= \int \int_0^1 E[Y_{1t}|X_t = x_t, V = v] f_{X_t}(x_t) dv dx_t, \\
E[Y_{0t}] &= \int \int_0^1 E[Y_{0t}|X_t = x_t, V = v] f_{X_t}(x_t) dv dx_t, \\
E[Y_{1t}|S = 1] &= \int \int_0^1 E[Y_{1t}|X_t = x_t, V = v] \frac{1 - F_{P|X}(v|x_t)}{\Pr(S = 1)} f_{X_t}(x_t) dv dx_t, \\
E[Y_{1t}|S = 0] &= \int \int_0^1 E[Y_{1t}|X_t = x_t, V = v] \frac{F_{P|X}(v|x_t)}{\Pr(S = 0)} f_{X_t}(x_t) dv dx_t, \\
E[Y_{0t}|S = 1] &= \int \int_0^1 E[Y_{0t}|X_t = x_t, V = v] \frac{1 - F_{P|X}(v|x_t)}{\Pr(S = 1)} f_{X_t}(x_t) dv dx_t,
\end{aligned} \tag{5}$$

and

$$E[Y_{0t}|S = 0] = \int \int_0^1 E[Y_{0t}|X_t = x_t, V = v] \frac{F_{P|X}(v|x_t)}{\Pr(S = 0)} f_{X_t}(x_t) dv dx_t.$$

2.3 Counterfactual Wage Distributions

In this section we describe estimators of $F_{Y_{1t}}(\cdot)$, $F_{Y_{0t}}(\cdot)$, $F_{Y_{1t}|S=1}(\cdot|S = 1)$, $F_{Y_{1t}|S=0}(\cdot|S = 0)$, $F_{Y_{0t}|S=1}(\cdot|S = 1)$, and $F_{Y_{0t}|S=0}(\cdot|S = 0)$, where $F_{Y_{1t}}(\cdot)$ is the c.d.f. of Y_{1t} , $F_{Y_{1t}|S=1}(\cdot|S = 1)$ is the c.d.f. of Y_{1t} conditional on $S = 1$, and other expressions are understood likewise. As in the previous section, our estimation strategy consists of two steps. In the first stage, we construct estimators of $F_{Y_{1t}|X_t, V}(y_{1t}|x_t, v)$ and $F_{Y_{0t}|X_t, V}(y_{0t}|x_t, v)$. In the second stage, the first stage estimators are integrated out with some suitable weights similar to (5). We use counterfactual wage distributions of Y_{1t} and Y_{0t} , namely $F_{Y_{1t}}(\cdot)$ and $F_{Y_{0t}}(\cdot)$, to study the evolution of inequality in the economy accounting for self-selection into schooling. As in the previous section, $F_{Y_{1t}|S=1}(\cdot|S = 1)$ and $F_{Y_{0t}|S=0}(\cdot|S = 0)$ can also be estimated directly by taking sample analogs of observed college and high school wages. Again, this provides some goodness-of-fit check of our model.

We now describe our estimators of $F_{Y_{1t}|X_t, V}(y_{1t}|x_t, v)$ and $F_{Y_{0t}|X_t, V}(y_{0t}|x_t, v)$. Note that these

⁸Notice that our estimation strategy depends on the assumption that $E[Y_{1t}|X_t = x_t, V = v]$ and $[Y_{0t}|X_t = x_t, V = v]$ can be estimated for any values of $v, 0 < v < 1$. That is, we need to assume that we observe full support for P for both college and high school wages.

densities can be obtained by location shifts from $F_{U_{1t}|V}(u_{1t}|v)$ and $F_{U_{0t}|V}(u_{0t}|v)$, that is

$$F_{Y_{1t}|X_t,V}(y_{1t}|x_t, v) = F_{U_{1t}|V}(y_{1t} - \mu_{1t}(x_t, \beta_{1t})|v) \quad \text{and}$$

$$F_{Y_{0t}|X_t,V}(y_{0t}|x_t, v) = F_{U_{0t}|V}(y_{0t} - \mu_{0t}(x_t, \beta_{0t})|v).$$

Therefore, once we estimate $\mu_{1t}(x_t, \beta_{1t})$ and $\mu_{0t}(x_t, \beta_{0t})$, we only need to estimate $F_{U_{1t}|V}(u_{1t}|v)$ and $F_{U_{0t}|V}(u_{0t}|v)$. As in (3) and (4), an application of Lemma 1 of Carneiro and Lee (2004) yields the following relationships

$$F_{U_{1t}|V}(u_{1t}|p) = F_{U_{1t}|P,S=1}(u_{1t}|p, S = 1) + p \frac{\partial}{\partial p} F_{U_{1t}|P,S=1}(u_{1t}|p, S = 1) \quad \text{and} \quad (6)$$

$$F_{U_{0t}|V}(u_{0t}|p) = F_{U_{0t}|P,S=0}(u_{0t}|p, S = 0) - (1 - p) \frac{\partial}{\partial p} F_{U_{0t}|P,S=0}(u_{0t}|p, S = 0). \quad (7)$$

Sample analogs of the right-hand sides of equations (6) and (7) can be obtained by some suitable nonparametric estimators. See Carneiro and Lee (2004) for details. Finally, we note that given estimators of c.d.f.'s, it is straightforward to obtain estimators of corresponding quantiles by inverting corresponding c.d.f.'s.

3 Revisiting the Recent Rise in the Returns to Schooling

In a recent paper, Card and Lemieux (2001) show that the rise in the return to schooling since the 1980s in the US (and in the UK and Canada) is mostly a phenomenon affecting young workers. In figure 1, we display a version of their main finding using Census data. The bottom two lines of the figure (labeled *ret*) are the OLS measures of the returns to college for white males by age group in two different Censuses: 1980 and 2000.⁹ The top four lines correspond to average log hourly wages in college (*col*) and high school (*hs*) deflated to 1992 dollars for each age group in each census. The age - return to schooling schedule for each year is just the difference between the college and high school line for each year. We add 1.5 to the returns to schooling at each age for scaling reasons. As emphasized by Card and Lemieux (2001), the increase in the returns to schooling

⁹We group individuals in the following age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, 60-65. We consider two schooling categories: the one we call high school (*hs* in the figure) consists of high school dropouts and high school graduates, while the one we call college (*col* in the figure) consists of individual with at least some college. Our income measure is the hourly wage computed by dividing the annual salary income by weeks worked per year and usual hours worked per week. We use the 1% extracts of the Census provided by IPUMS.

between 1980 and 2000 is especially large for young individuals. They argue that the main reason in such phenomenon is the slowdown in the growth of educational attainment for cohorts born after 1950. Figure 2 illustrates this claim using census data. It plots educational attainment by age group for the 1980 and 2000 censuses. Taking the 1980 census, educational attainment grows substantially until we reach the 1950 cohort, whose individuals are aged 30 by 1980. The 2000 census shows a similar story, although it accentuates the peak in college attendance due to the vietnam draft also documented in many other papers (Angrist??). If the demand for skilled workers grows at a constant rate and if workers of different ages are imperfect substitutes then a slowdown in the rate of increase in the supply of skilled workers will lead to an increase in their relative price.

Figure 1 also shows that most of the increase in the returns to schooling between 1980 and 2000 is due to a decrease in the average wages of individuals in the high school group. The average age-wage profile in the college group is basically the same in 1980 and 2000. The prediction of Card and Lemieux's model for the evolution of the levels of average wages depends not on the changes in relative demand and supply of skills, but on the changes in the absolute demand and supply of labor for each skill group, which in turn depends, among other things, on the size of each cohort. But what's more relevant for the topic of our paper is that the prediction of Card and Lemieux's model for the evolution of inequality depends on how we account for heterogeneity and selection. Their model is essentially a representative agent model within each age group, where all individuals within each of these categories face the same price for their skills. In such a model, whatever changes we observe in skill prices they should translate into location shifts in the distribution of wages for each age-skill group, unless institutions such as minimum wages or unions significantly alter the reaction of the wage distribution to changes in prices (see, for example, DiNardo, Fortin and Lemieux, 1996, and Lee, 2003). However, if individuals are heterogeneous and self-select into schooling levels then changes in educational attainment not only produce fundamental changes in prices but they also change the (ability) composition of different education groups (as emphasized by, for example, Acemoglu, 2003, and Juhn, Kim and Vella, 2005). Figures 3, 4 and 5 show how the distribution of log hourly wages within each age-skill group changed from 1980 to 2000. The top two lines of figure 3 correspond to the 10th percentile of the college wage distribution for each age group and census year. The middle two lines correspond to the 10th percentile of the high school wage distribution. The bottom two lines are the difference between the 10th percentile of the college

and high school wage distributions for each age and census year.¹⁰ Figures 4 and 5 present the corresponding graphs for the 50th and 90th percentiles of the wage distributions. It is clear that any potential change in skill prices driven by supply and demand forces has not translated into simple location shifts in the distribution of wages. We observe different changes in different percentiles for different skill groups, although the differences between the college and high school wage distributions grow especially for individuals at young ages for the three quantiles we consider (which translates into a rise in the least squares measure of the returns to schooling especially for individuals of young ages).¹¹

An alternative framework where there are heterogeneous individuals, multiple skills and purposive sorting of individuals across schooling levels may be helpful in the understanding of the patterns we just documented. For example, it is well known that real wages have fallen since the 1980s in the middle and lower tail of the wage distribution in the US while they have risen at the top (see, for example, Murphy and Welch, 1999). However, among college graduates, wages have not fallen as much for individuals at young ages. Perhaps they would have fallen much more if college enrolment had increased for this group of individuals leading to a decline in the average quality of the college graduate. For previous generations, the decline in real wages may be due both to a decline in quality due to increasing schooling attainment and/or a fundamental decline in college wages at the bottom of the distribution. Furthermore, it is even possible that the price of college skills increase between 1980 and 2000 leading to an increase in wages in the top percentiles of the wage distribution, which is not as affected by changes in the distribution of quality of college graduates as lower percentiles of the wage distribution.¹²

How would changes in composition affect the high school wage distribution? First, recall that in our current work we aggregate high school graduates and high school dropouts in the same group. In the absence of changes in skill prices, changes in the high school wage distribution are caused by

¹⁰This is what is usually called a quantile treatment effect, although this is a misleading term since in general this is not a return to schooling for anyone in the economy even in the absence of selection, unless we assume that individuals are equally ranked in the college and high school wage distributions (see Heckman, Smith and Clemens, 1996).

¹¹Changes in the composition of the non-employed population can also lead to changes in inequality, as in Blundell, Reed and Stoker (2003) or Blundell, Gosling, Ichimura and Meghir (2005). However there are no substantial changes in the age profile of non-employment between these two censuses, as shown in figure A1 in the appendix. A relevant sidepoint of this figure is the very steep negative gradient in the age-employment profile. Carneiro (2005) analyses the impact of differential non-employment by age for the estimation of empirical age-wage profiles.

¹²It is also possible that there are multiple skills and those individuals at the top of the wage distribution also have the skill whose price increased the most.

individuals shifting from being dropouts to being high school graduates, and individuals leaving the high school group and joining the college group. For now, we ignore the first channel and concentrate on the second. High school dropouts are probably at the bottom of the wage distribution and will not be the ones moving into college as educational attainment increases. Therefore, we expect that those at individuals at the middle and top of the high school wage distribution to be at the margin between enrolling in college or not. If those individuals who are the margin between going to college or not do not come from the top of the high school wage distribution (as in Willis and Rosen, 1979, Carneiro, Heckman and Vytlačil, 2003, and Carneiro and Lee, 2004), then a decrease in college attendance may also lead to a decrease in high school wages, because they are high school graduates of lower than average quality. Therefore, if college attainment were declining the evolution of the high school wage distribution would not be surprising.

In the data we do not observe a decrease in college attainment. What we observe is a slowdown in its growth. The accurate way to rephrase our tentative analysis would be the following: the decline in the quality of the college population due to the secular increase in college attainment more interacted with a rising demand for skill to produce the observed patterns of wage inequality. For recent generations, this decline in quality has stopped.

Even though this is highly tentative reasoning, it is also suggestive of the power and importance of a more flexible framework for the study of inequality. A more developed version of the work sketched in this section considering more skill groups is applied to several censuses and the Current Population Survey (CPS) for the US and to the General Household Survey (GHS) for the UK in Carneiro and Lee (2005). In the next section we show additional and more direct evidence of the importance of accounting for heterogeneity and selection into schooling for analyzing the evolution of wage inequality.

4 Empirical Results

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 (see, for example, Katz and Murphy, 1992, Murphy and Welch, 1992, Juhn, Murphy and Pierce, 1993, DiNardo, Fortin and Lemieux, 1998, Katz and Autor, 1999). The sample sizes in the CPS are large and the data is collected annually. However

the relative lack of detail on each individual in the CPS does not allow us to properly model the process of selection into schooling, which is needed for an analysis of the role of sorting into schooling for the study of inequality. For this purpose, the NLSY is much better since the individual data available in this survey is much richer than in the CPS, especially in terms of the variables that determine the educational attainment of each person.¹³ In the NLSY we have detailed information on cognitive ability and family background, which are important determinants of both schooling and labor market outcomes. Furthermore, for most respondents in the NLSY we know the location of their residence during the adolescent years. School and labor market characteristics in different areas of residence of adolescent NLSY respondents will be used as instrumental variables for schooling, allowing us to identify the role of sorting in the labor market in determining inequality.¹⁴ 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 work available on request from the authors we show that we can replicate the evolution of inequality in the CPS in the 1990s for the cohorts surveyed by the NLSY using NLSY data.

The reason we choose to start our analysis in the 1990s and not before is because NLSY respondents were very young in the 1980s and would have been recent entrants into the labor market at that time. Even though many of our respondents are still relatively young during the period of 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. Furthermore, inequality in the 1990s is relatively much less analysed than inequality in earlier periods. Finally, in this paper we focus on white males only.

Our sample consists of white males born between 1957 and 1964. We use the hourly wage measure created by the NLSY. Unemployment is not an important concern in this population for the time period of our study.¹⁵ Furthermore, in order to minimize measurement error and further reduce concerns with selective unemployment we use as our wage measure in each year a 5 year average of all non-missing wages reported in the five year interval centered in our year of interest.

¹³For a description of the NLSY see BLS (2003). The data we use in this paper is described in detail in the appendix.

¹⁴The NLSY has another important advantage: it is a longitudinal study. However, because in this version of the paper we are still not making use of the longitudinal structure of the NLSY we do not emphasize this advantage yet.

¹⁵In the appendix we report information on missing wages for our sample for each year of the analysis.

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 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. However, our selection model only considers selection on unobservables for two broad groups: high school plus dropouts vs. some college plus college. The main reason for doing this aggregation is the lack of the credible exclusion restrictions needed to estimate a model with multiple levels of schooling (see Heckman and Vytlačil, 2004). Furthermore, this aggregation is very common in the literature on wage inequality where often only two schooling groups are considered.

We now turn to the exact specification of the selection model we estimate. The X vector in the log wage equations consists of experience, schooling-adjusted score on the Armed Forces Qualifying Test (or AFQT, a measure of cognitive ability)¹⁶, number of siblings, mother's years of schooling, father's years of schooling, and the local unemployment rate in the current area 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. In addition, we include a dummy variable for being a high school dropout, another dummy variable for being an college attendee without a college degree, and interactions of these variables with quadratic polynomials in experience and AFQT.

The Z vector in the school choice equation consists of AFQT, number of siblings, mother's education, father's education, distance to college at age 14, average tuition in the county the individual lived in at age 17, and the unemployment rate in the state of residence in 1979. The variables that we exclude from the outcome equations are distance to college, tuition, and local unemployment rate. The AFQT and family background variables 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 that we estimate all variables enter with a linear and a quadratic term and are interacted

¹⁶We adjust the AFQT score by the amount of schooling each individual has at the time they take the test (see Hansen, Heckman and Mullen, 2003, and the appendix to this paper).

with each other.

The exclusion restrictions we use are based on the geographic location of individuals at age 17, conditional on family background variables and measured ability. If the decision of going to college and the location decision are correlated then our instruments are not valid. For example, individuals who are more likely to enroll in college may choose to locate in 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 (2000) 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 in the outcome equation. Tuition was used by Kane and Rouse (1995). Average tuition in the county of residence may also be a problematic instrument if it is correlated with average college quality in the county. Our assumption is that the ability and family background variables that we have 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, Cameron and Taber (2004) 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. They use county level local labor market variables while we use state level variables. In the outcome equations we estimate we include the local unemployment rate in the year in which wages are measured.

Sample statistics are presented in table 1. Individuals who attend college have on average higher wages than those who do not attend college. They also have higher levels of cognitive ability, more educated parents, live nearer to colleges, in counties with lower average tuition and in states with higher unemployment rates at age 17 than those individuals who never enrolled in college. Notice that even though we are following a cohort of individuals well into their adult years the college enrolment rate in our sample increases substantially over time. This is not a particular feature of this dataset. Similar patterns can be found by following similar cohorts in the CPS. Educational

attainment only seems to stabilize for a given cohort around age 45.

We use a logit model for schooling choice. That is, $S = 1$ if $P(Z) > V$, where $P(z) = L[\mu_S(z)]$ and $L(u) = \exp(u)/[1 + \exp(u)]$. 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. 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. We estimate one logit model for each year. Average derivatives are presented in table 2. Ability and family background are strong predictors of college attendance. Tuition is also an important determinant of enrolment 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 enrolment of some specific groups of individuals, although not for all individuals.

The identification of the objects of interest was discussed in the previous section. The exact procedure which we use to estimate this model is described in detail in Carneiro and Lee (2004). For each year we estimate $f(Y_1|X, V)$ and $f(Y_0|X, V)$ and then weight these objects with appropriate weights to construct the counterfactuals of interest, as described in the previous section. 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 for which we have individuals both in the college and the high school group, and for values of P (and V) for which we have individuals both in the college and high school group. Figure 6 shows the support of the data for 1994, a representative year in our sample.¹⁷ The top two figures refer to P and the bottom two figures refer to AFQT. AFQT is only one of the variables on the X vector on which we condition on, but it is the most important one. 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.

We start by presenting our estimates of $E(Y_1|X, V)$, $E(Y_0|X, V)$, $f(Y_1|X, V)$ and $f(Y_0|X, V)$ (where $f(\cdot)$ denotes p.d.f.) for 1994, the second year of our sample. $E(Y_1 - Y_0|X, V)$ is the marginal treatment effect of Heckman and Vytlačil (2000, 2004) and it is a natural way to look at heterogeneity in returns. In figure 7 we graph $E(Y_1|AFQT, \text{experience} = 10, V = 0.5)$, $E(Y_0|AFQT, \text{experience} = 10, V = 0.5)$ and $E(Y_1 - Y_0|AFQT, \text{experience} = 10, V = 0.5)$, as functions of AFQT. We fix years of experience

¹⁷Other years are available from the authors on request. This figure changes very little across years.

at 10 to abstract from life-cycle effects, V at its median value and remaining variables in X at 3 siblings, 12 years of mother's and father's education, 7% of local unemployment rate, and zeros for high school dropout and some college dummies. In figure 8 we graph $E(Y_1|AFQT = 0, \text{experience} = 10, V)$, $E(Y_0|AFQT = 0, \text{experience} = 10, V)$ and $E(Y_1 - Y_0|AFQT = 0, \text{experience} = 10, V)$. 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. We focus on AFQT and V because they are both strong determinants of selection and they are strongly correlated with wages. The effect of AFQT on wages is very strong 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 the smaller the likelihood that an individual enrolls in college. The results here confirm the findings in Carneiro, Heckman and Vytlačil (2003) and Carneiro and Lee (2004). There are strong selection effects both in levels and in returns, and the main story in the labor market seems to be one of both comparative and absolute advantage (as in Willis and Rosen, 1979): the individuals who have the highest potential wages in the college sector decide to enrol in college, and those with the highest potential in the high school sector do not enrol in college. Individuals with high college wages would have low high school wages if they did not go to college, and vice versa. These estimates imply that we cannot have a single skill model of the labor market. The amount of heterogeneity agents select on¹⁸ is very large. Similar patterns are found for other years, although with some differences. In figure A2 in the appendix we plot these curves for all the years we analyze.

Y_1 and Y_0 are just a convolution of two random variables ($\mu_1(X)$ and U_1 , and $\mu_0(X)$ and U_0 , respectively)¹⁹, so they can be written as a function of the conditional densities of these random variables. Since we assume that (U_1, V) and (U_0, V) are independent of X , implying that we do not allow for any heteroskedasticity as a function of X , it suffices to focus on $f(U_1|V)$ and $f(U_0|V)$. In figure 9 we graph the 25th, 50th and 75th percentiles of these two conditional densities as functions of V . In this figure, U_1 and U_0 are normalized to have mean zero. All the quantiles of U_1 decrease with V , confirming what we already knew from the study of the mean: individuals less likely to enrol in college (in terms of V) have lower wages in the college sector. However, notice that there still is a

¹⁸*Ex ante* heterogeneity, emphasized in Carneiro, Hansen and Heckman (2003) and Cunha, Heckman and Navarro (2004, 2005).

¹⁹We define $\mu_1(X)$ and $\mu_0(X)$ as random variables as well.

considerable amount of dispersion in wages. However, this dispersion does not change substantially with V . For U_0 , the way its quantiles vary with V parallels the movements of the mean, but the dispersion of U_0 is higher for higher values of V . Carneiro, Hansen and Heckman (2003) and Cunha, Heckman and Navarro (2004, 2005) interpret $E(U_1|V)$ and $E(U_0|V)$ as *ex-ante* heterogeneity and $f(U_1|V)$ and $f(U_0|V)$ as uncertainty. In their framework, uncertainty in college wages does not vary substantially across different individuals, while uncertainty in high school wages is higher for individuals who are less likely to enrol in college. This heteroskedasticity plays an interesting role in our analysis of the effect of increases in educational attainment on wage inequality. For example, when educational attainment increases, the density of (X, V) changes within the high school (and college) group. This has an effect in the observed distribution of wages of those individuals who enrolled in college both because $E(Y_0|X, V)$ varies with (X, V) , but also because the shape of $f[U_0 - E(U_0|V)|V]$ varies with V . In particular, for a fixed X , if those individuals leaving the high school sector have a V which is higher than the average in their sector, average wages in high school will tend to increase. Overall dispersion may decrease since the distribution of V becomes more concentrated in high values, but it may also increase since the dispersion of U_0 increases with V .

We estimate the model for five years: 1992, 1994, 1996, 1998 and 2000. Figure 10 shows estimates of $E(Y_{1t}|S_t = 1)$ and $E(Y_{0t}|S_t = 0)$ for each year using our model, and compares these with the actual data. Even though the computation of these estimates involves a multiple step procedure, the model estimates and the data match almost perfectly, which suggests we are using an accurate procedure (as a result, the model also fits relatively well the level and evolution of between group inequality the way it is usually measured: $E(Y_{1t}|S_t = 1) - E(Y_{0t}|S_t = 0)$). In figure 11, We then compare the evolution of two specific counterfactuals of interest, $E(Y_{1t})$ and $E(Y_{0t})$, to simple estimates of $E(Y_{1t}|S_t = 1)$ and $E(Y_{0t}|S_t = 0)$. If the distribution of X and V is fixed in the population, from the evolution of $E(Y_{1t})$ and $E(Y_{0t})$ we can infer changes in skill prices, since these objects are purged of composition effects and selection bias. The growth of $E(Y_{1t}|S_t = 1)$ is well below the growth in $E(Y_{1t})$, while selection bias does not seem to generate large differences between the growth of $E(Y_{0t}|S_t = 0)$ and $E(Y_{0t})$. If, for example, increasing demand for college graduates is increasing their wages, we would underestimate the increase in price by looking at $E(Y_{1t}|S_t = 1)$ instead of $E(Y_{1t})$.

Figure 12 shows the evolution of the OLS estimate of the return to education (a standard mea-

sure of between group wage inequality) and three different average returns: the average return for individuals enrolled in college (TT), the average return for individuals not enrolled in college (TUT) and the average return for a random person in the economy (ATE). While we cannot directly link the evolution of the OLS estimate with a change in the price of skill, we can do that for TT, TUT and ATE. The OLS estimate smoothly increases over time. TUT and ATE increase over time, but at different rates, while TT is roughly stable until the last period. Which measure (if any) we want to use to describe the evolution of between wage inequality or changes in the price of skill, depends on the specific question of interest (see Heckman and Vytlačil, 2004, and Carneiro, Heckman and Vytlačil, 2003).

Figures 13 and 14 present estimated and actual quantiles of $f(Y_1|S=1)$ and $f(Y_0|S=0)$. The model fits the data very well. There are, however, slight discrepancies, especially in the lower tails of both these distributions. Fitting the tails probably requires a much larger dataset than the one we have available. Therefore, in our analysis of 75-25 wage differentials may be more reliable than our analysis 90-10 differentials (which are shown in the appendix). Still, the fit of the differentials of the 75th and 25th percentiles of the wage distributions for these two schooling groups is worse than the fit of their levels, as shown in figures 15 and 16.

Within group 75-25 wage differentials are shown in figures 17 and 18. Both figures present the actual (data) evolution of these wage differentials for each group ($f(Y_1|S=1)$ and $f(Y_0|S=0)$) and a counterfactual evolution in wage differential that we would observe if every individual in the economy chose the same education level or if individuals were randomly assigned across schooling levels ($f(Y_1)$ and $f(Y_0)$). The counterfactual evolution of inequality is purged of selection and therefore is a better way of identifying changes in skill prices and their impact on the wage distribution than just looking at the evolution of the wage distribution for each group.²⁰ Figure 15 shows that, among college educated individuals, the 75-25 differential increased slightly from 1992 to 2000. This was due to increased inequality at the top of the wage distribution: the 75-50 differential increases over this period while the 50-25 differential stays constant. The counterfactual and actual evolution of college wage inequality are very similar. This is surprising given that we found large differences in the evolution of the mean. In the appendix we show that for estimates of the 90-10 differential the

²⁰Many other counterfactuals could be defined and analyzed instead, depending on the question of interest (see Heckman and Vytlačil, 2004, Carneiro, Hansen and Heckman, 2003). We view $f(Y_1)$ and $f(Y_0)$ as the most useful counterfactuals to study price changes.

story is somewhat different: the measured increase understates the counterfactual increase in college wage inequality. Figure 16 shows that, among high school graduates, the sample 75-25 differential also increased from 1992 to 2000. This change is driven solely by movement in the 50-25 differential, since the 75-50 differential is roughly constant over time. However, when we look at counterfactual wage differentials, the increase in inequality is considerably larger, especially at the top of the wage distribution.²¹

In summary, accounting for selection clearly changes the perception we have about movements in the underlying prices of skill in the economy. For example, while the OLS estimate of the return to schooling increases by 2% over our period of analysis, the average return to schooling in the economy ($E(Y_1 - Y_0)$) increases by 8%. $E(Y_{1t}|S_t = 1)$ increases by 2% while $E(Y_{1t})$ increases by more than 4%. The 90-10 differential increases by 1% when we look at $f(Y_{1t}|S_t = 1)$ but it increases by 2% when we look at $f(Y_{1t})$. Neglecting selection leads to underestimate the growth in the price of college skills, which lead to an increase in potential college wages, average returns to schooling and within group inequality.²²

We can now attempt to interpret the change in inequality as a function of the variables of our model. One alternative would be, for example, to follow the procedure of Juhn, Murphy and Pierce (1993) in the way they decompose the increase in inequality as a function of observable and unobservable variables. We adopt a similar idea but use a different approach. Figure 19 plots the counterfactual and observed evolution of 90-10, 90-50 and 50-10 differentials of the distribution of college wages. By counterfactual distribution we mean $f(Y_1)$ and by observed distribution we mean $f(Y_1|S = 1)$. 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 (squares) and a third one that plots the evolution of inequality if the distribution of X were kept at its 1992 values (X). In the appendix we explain the exact procedure used. From the three left panels it is clear that most of the increase in the evolution of counterfactual 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

²¹The standard way to examine changes in within group inequality is not even the way described in this paper. Often researchers take the residuals of a regression of log wages on education and experience and call the dispersion of these residuals within group inequality. This is a very restrictive procedure. The distributions of log wages for high school and college are clearly different and should be looked at separately.

²²At this stage we did not yet spell out a concrete model of pricing of skills, but, as Acemoglu (1991) emphasizes, it has to be a model of multiple skills to allow for an increase in the dispersion of log wage inequality.

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, but may have induced an increase in inequality at the bottom of the college wage distribution around 1996. If we just did this decomposition (or other types of decomposition, such as the ones in Juhn, Murphy and Pierce, 1993, DiNardo, Fortin and Lemieux, 1998, or Autor, Katz and Kearney, 2005) using the observed variation in inequality, and therefore using variation that does not correspond only to changes in prices, we get different answers. At the top of the wage distribution, the increase in observed inequality 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 its movements are explained by the growth of inequality due to both X and U_1 .²³

Figure 20 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 figure 16, 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 . Changes in $\mu_0(X)$ do not seem to translate in any significant change in high school wage inequality. This may be due to the fact that AFQT, the main determinant of wages in our X vector, plays a bigger role in college than in high school.

Finally, we do the converse but equally useful exercise: what would happen to inequality if prices were fixed but the composition of each education group changed over time? We estimate the model and simulate how the level of inequality in the economy would change if college enrolment rates varied from small to large. One way to think about this experiment is the following: if the government reduced the costs of college attendance more and more individuals would enrol in college and the composition of the college group would change; conversely, if the government increased the costs of college attendance less and less individuals would enrol in college. We simulate exogenous shifts in the cost of schooling by changing the intercept of our selection equation. We present the results of the simulation on figure 21. The parameters of our model come from estimating our regression using data for 1992, but we could have picked any other year to perform a similar exercise. We do not simulate settings where the average college enrolment rate is either very close to 0 or very close to 1.

²³As usual, we could do similar types of decomposition using any other year as our base year. In the appendix we show that using 2000 as our base year instead of 1992 does not change our basic conclusions.

Instead, we restrict ourselves to cases where the average college enrolment rate is between 0.34 and 0.66 (the mean in the data is close to 0.5). An important concern with our simulations is that our 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. Hopefully, by restricting ourselves to a smaller (but still large) range of values than the full possible support of changes, our simulation results will be more realistic.

Figure 21 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 parameter rises sharply as the proportion of individuals attending college decreases. The reason behind this is that the population of college graduates becomes more selective and of better quality as tuition increases, and therefore their average wages increase. Simultaneously, the population of high school graduates becomes of lower quality and average wages decrease. 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 distribution. 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 and median high school wages stay flat as the proportion of individuals with a college education increases. The overall median wage increases as more individuals enter the college group. Interestingly, we observe the same behaviour in all (graphed) percentiles of the high school, college and overall distribution of wages. Furthermore, the observed (conditional on $S=1$) 90-10 college wage differential increases, but the 90-10 high school wage differential stays flat. 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. As a result, overall inequality could increase or decrease. In our sample, the overall 90-10 wage differential is increasing with college enrolment for all the values of college enrolment that we consider: increases in educational attainment lead to increases in inequality.

5 Conclusion

This paper shows that accounting for selection produces important quantitative and qualitative changes in the main patterns of inequality. It is important to *do* incorporate selection in the study of inequality for at least two reasons. First, in order to understand what fundamental economic forces that shape inequality, we need to obtain accurate measures of changes in the underlying prices of skill in the economy. Second, to estimate the effect of different policies on wage inequality we need to know how individuals sort into different levels of schooling. We find that accounting for selection produces large differences in our estimates of wage changes over time, in our decompositions of changes in overall inequality into different factors, and in our simulations of the effects of education policy on wage inequality.

References

- [1] Abadie, A. (2002), “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models,” *Journal of the American Statistical Association*, 97: 284-292.
- [2] 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.
- [3] 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.
- [4] Angrist, J. and A. Krueger (1991), “Does Compulsory School Attendance Affect Schooling and Earnings?”, *The Quarterly Journal of Economics*, Vol. 106, No. 4., pp. 979-1014.
- [5] 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.
- [6] Black, D. and J. Smith (2004), “How Robust is the Evidence on the Effects of College Quality? Evidence from Matching”, forthcoming in *Journal of Econometrics*.
- [7] Blundell, R., L. Pistaferri and I. Preston (2002), “Partial Insurance, Information and Consumption Dynamics”, IFS Working Paper W02/16.

- [8] Blundell, R., H. Reed and T. Stoker (2003), "Interpreting Aggregate Wage Growth: The Role of Labour Market Participation", *American Economic Review*, Vol. 93, No. 4, September, 1114-1131.
- [9] Bureau of Labor Statistics (2001), *NLS Handbook*, 2001, Washington, DC: US
- [10] Card, D. (1993), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling", NBER wp 4483.
- [11] 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.
- [12] Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5): 1127-60.
- [13] Card, D. and T. Lemieux (2001), "Can Falling Supply Explain the Rising Return to College For Younger Men? A Cohort Based Analysis," *The Quarterly Journal of Economics*.
- [14] Cameron, S. and J. Heckman (2001). "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109:455-499.
- [15] Cameron, S. and C. Taber (2004), "Estimation of Educational Borrowing Constraints Using Returns to Schooling", *The Journal of Political Economy*, Volume 112, Number 1, part 1.
- [16] Carneiro, P. (2003a), "Heterogeneity in the Returns to Schooling – Implications for Policy Evaluation." Ph.D. dissertation, University of Chicago.
- [17] Carneiro, P. (2003b), "Methods for Evaluating the Impact of Education Policies when the Return to Education Varies Across Individuals", *Cuadernos de Economia*, Ano 40, No. 121, pp. 516-529.
- [18] Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post-secondary Schooling," *Economic Journal* 112(482): 705-34.
- [19] Carneiro, P. and J. Heckman (2004), "Empirical Estimates of the Returns to Schooling", working paper.

- [20] 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.
- [21] 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.
- [22] Carneiro, P and S. Lee (2004), "Comparative Advantage and Schooling", working paper, University College London.
- [23] Chandra, A. (2003), "Is the Convergence of the Racial Wage Gap Illusory?," NBER working paper 9476.
- [24] Chernozhukov, V. and C. Hansen (2004), "An IV model of Quantile Treatment Effects," forthcoming, *Econometrica*.
- [25] Chesher, A. D. (2003), "Identification in Nonseparable Models," *Econometrica*, 71: 1405-1441.
- [26] Cunha, F., J. Heckman and S. Navarro (2003), "Counterfactual Analysis of Inequality and Social Mobility", working paper, The University of Chicago.
- [27] Cunha, F., J. Heckman and S. Navarro (2004), "Separating Heterogeneity From Uncertainty In Income and Human Capital Dynamics", working paper, The University of Chicago.
- [28] Dale, S. and A. Krueger (1999), "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables", NBER wp 7322.
- [29] Das, M., W. K. Newey and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economics Studies*, 70: 33-58.
- [30] Deschenes, O. (2001)
- [31] DiNardo, Fortin and Lemieux (1996)
- [32] Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications* (London: Chapman & Hall).

- [33] Flavin, M., "The Adjustment of Consumption to Changing Expectations about Future Income," *Journal of Political Economy* 89(1981), 974-1009.
- [34] Florens, J., J. Heckman, C. Meghir and E. Vytlacil (2002), "Instrumental Variables, Local Instrumental Variables and Control Functions", CEMMAP working paper CWP15/02.
- [35] Hansen, K., K. Mullen and J. Heckman (2003), "The Effect of Schooling and Ability on Achievement Test Scores," forthcoming in *Journal of Econometrics*.
- [36] Heckman, J. and B. Honoré (1990), "The Empirical Content of the Roy Model", *Econometrica*, vol. 58, no. 5, pp. 1121-1149.
- [37] Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.
- [38] Heckman, J. H. Ichimura and P. Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *The Review of Economic Studies* 64:4, 605-654.
- [39] Heckman, J. H. Ichimura and P. Todd (1998), "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies* 65:2, 261-294.
- [40] 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.
- [41] Heckman, J. and J. Scheinkman (1986), "The Importance of Bundling in a Gorman-Lancaster Model of Earnings", *The Review of Economic Studies*, Vol. 54, No. 2., pp. 243-255.
- [42] Heckman, J. and G. Sedlacek (1985), "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *The Journal of Political Economy*, Vol. 93, No. 6., pp. 1077-1125.
- [43] Heckman, J. and G. Sedlacek (1990), "Self-Selection and the Distribution of Hourly Wages", *Journal of Labor Economics*, Vol. 8, No. 1, Part 2: Essays in Honor of Albert Rees, pp. S329-S363.

- [44] 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.
- [45] 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.
- [46] Heckman, J. and E. Vytlačil (2000), "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, 2000), 1-46.
- [47] Heckman, J. and E. Vytlačil (2001), "Policy Relevant Treatment Effects," *American Economic Review Papers and Proceedings*, 91(2): 107-111.
- [48] Heckman, J. and E. Vytlačil (2004), "Structural Equations, Treatment, Effects and Econometric Policy Evaluation," forthcoming in *Econometrica*.
- [49] Ichimura, H., and C. Taber (2000), Direct Estimation of Policy Impacts, Northwestern University and University College London, unpublished manuscript.
- [50] Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2):467-475.
- [51] Imbens, G. and W. Newey (2003), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity", working paper, MIT.
- [52] Imbens, G. and D. Rubin (1997), "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economics Studies*, 64: 555-574.
- [53] Juhn, C., K. Murphy and B. Pierce (1993), "Wage Inequality and the Rise in Returns to Skill", *The Journal of Political Economy*, Vol. 101, No. 3., pp. 410-442.
- [54] Kane, T. (1994), "College Entry by Blacks Since 1970: The Role of College Costs, Family Background and the Returns to Education," *Journal of Political Economy*, 102(5): 878-911.

- [55] Kane, T. and C. Rouse (1995), "Labor-Market Returns to Two- and Four-Year College", *The American Economic Review*, Vol. 85, No. 3., pp. 600-614.
- [56] Katz, L. and K. Murphy (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *The Quarterly Journal of Economics*, Vol. 107, No. 1. (Feb., 1992), pp. 35-78.
- [57] Linton, O., E. Maasoumi and Y.-J Whang (2003), "Consistent Testing for Stochastic Dominance under General Sampling Schemes," working paper, LSE.
- [58] Mandelbrot, B. (1962), "Paretian Distributions and Income Maximization", *The Quarterly Journal of Economics*, Vol. 76, No. 1, pp. 57-85.
- [59] McFadden, D. (1989), "Testing for Stochastic Dominance," in T. B. Fomby and T. K. Seo, (eds.), *Studies in the Economics of Uncertainty in Honor of Josef Hador*, (New York: Springer Verlag).
- [60] National Center for Education Statistics (2003), *Digest of Education Statistics*, 2002, US Department of Education.
- [61] Neal, D. (2004), "The Measured Black-White Wage Gap among Women Is Too Small", *The Journal of Political Economy*, Vol. 112, No. S1.
- [62] Roy, A. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3:135-146.
- [63] Sattinger, M. (1978), "Comparative Advantage in Individuals", *The Review of Economics and Statistics*, Vol. 60, No. 2., pp. 259-267.
- [64] Sattinger, M. (1993), "Assignment Models of the Distribution of Earnings", *Journal of Economic Literature*, Vol. 31, No. 2., pp. 831-880.
- [65] Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
- [66] Vytlačil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*.70(1): 331-41

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

Figure 1 - The Evolution of the Returns to College for Individuals with Different Ages

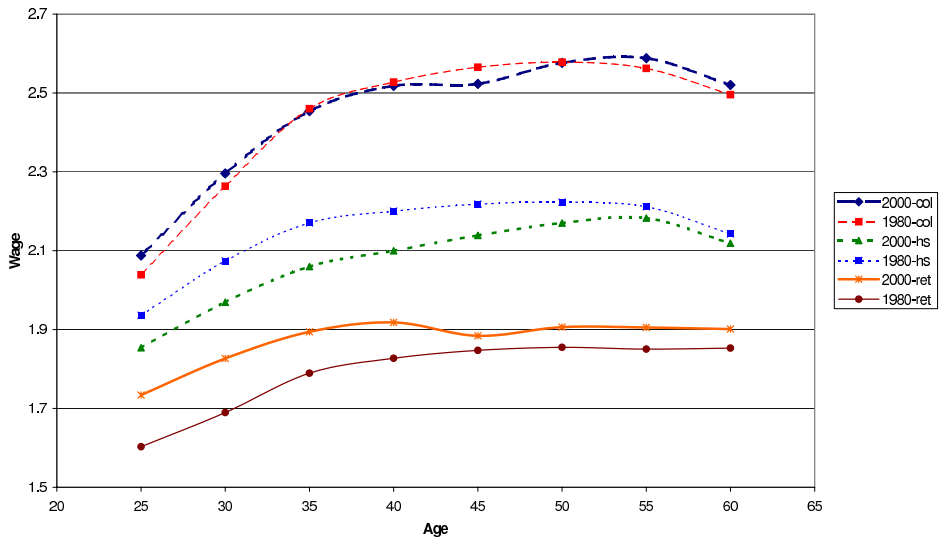


Figure 1.

Figure 2 - Percentage of Individuals with at least Some College Attendance

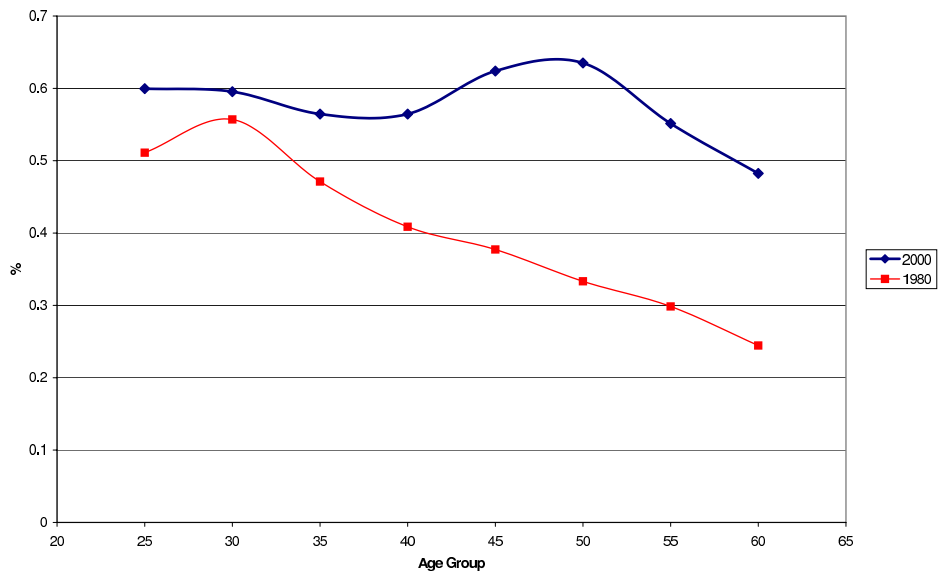


Figure 2.

Figure 3 - 10th Percentile of the Wage Distribution for Individuals with Different Ages

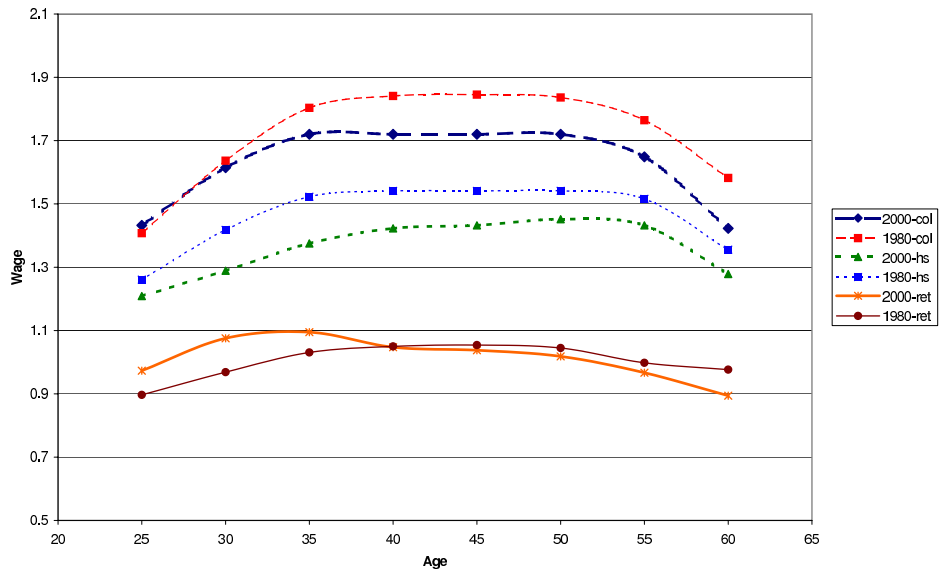


Figure 3.

Figure 4 - 50th Percentile of the Wage Distribution for Individuals with Different Ages

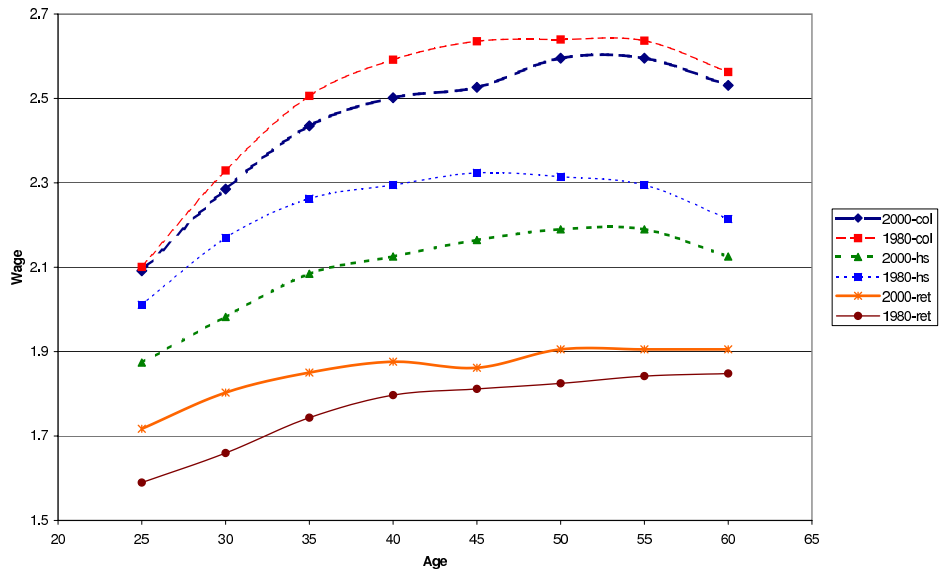


Figure 4.

Figure 5 - 90th Percentile of the Wage Distribution for Individuals with Different Ages

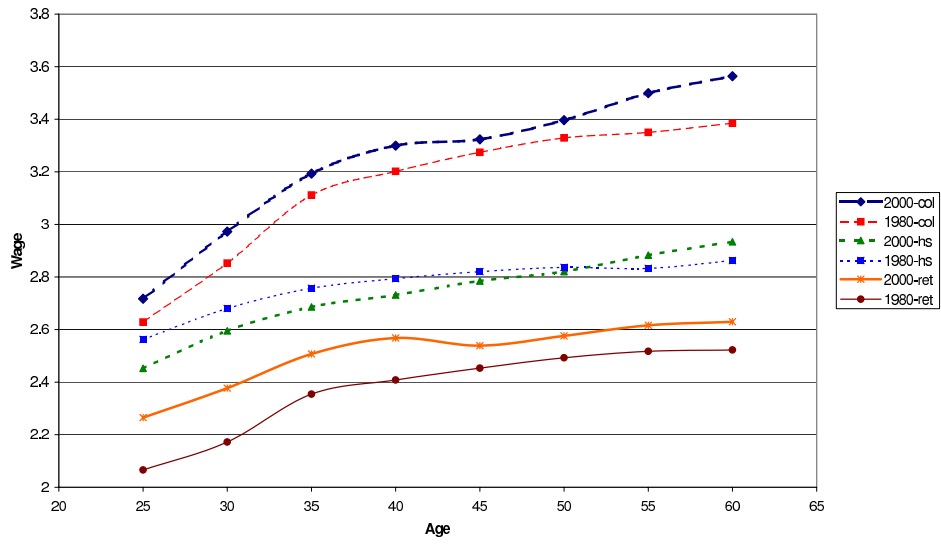


Figure 5.

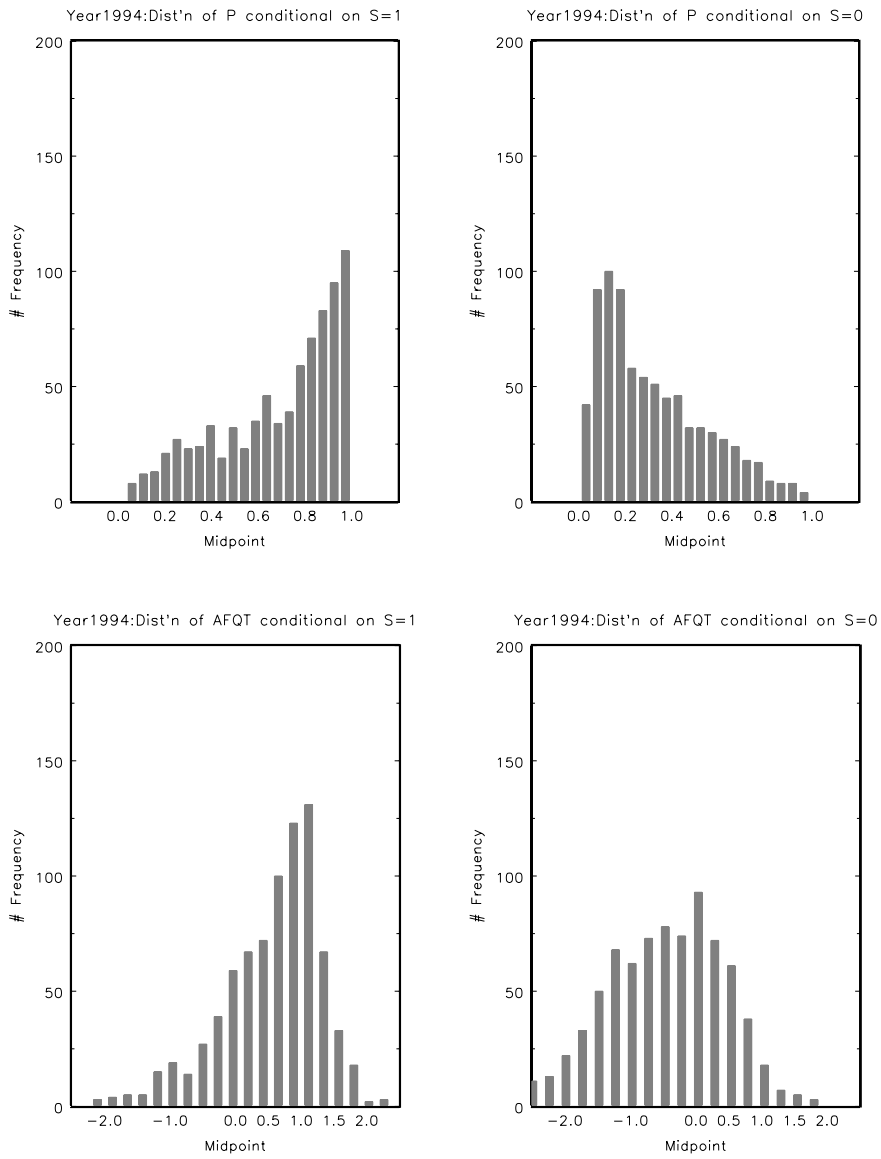


Figure 6.

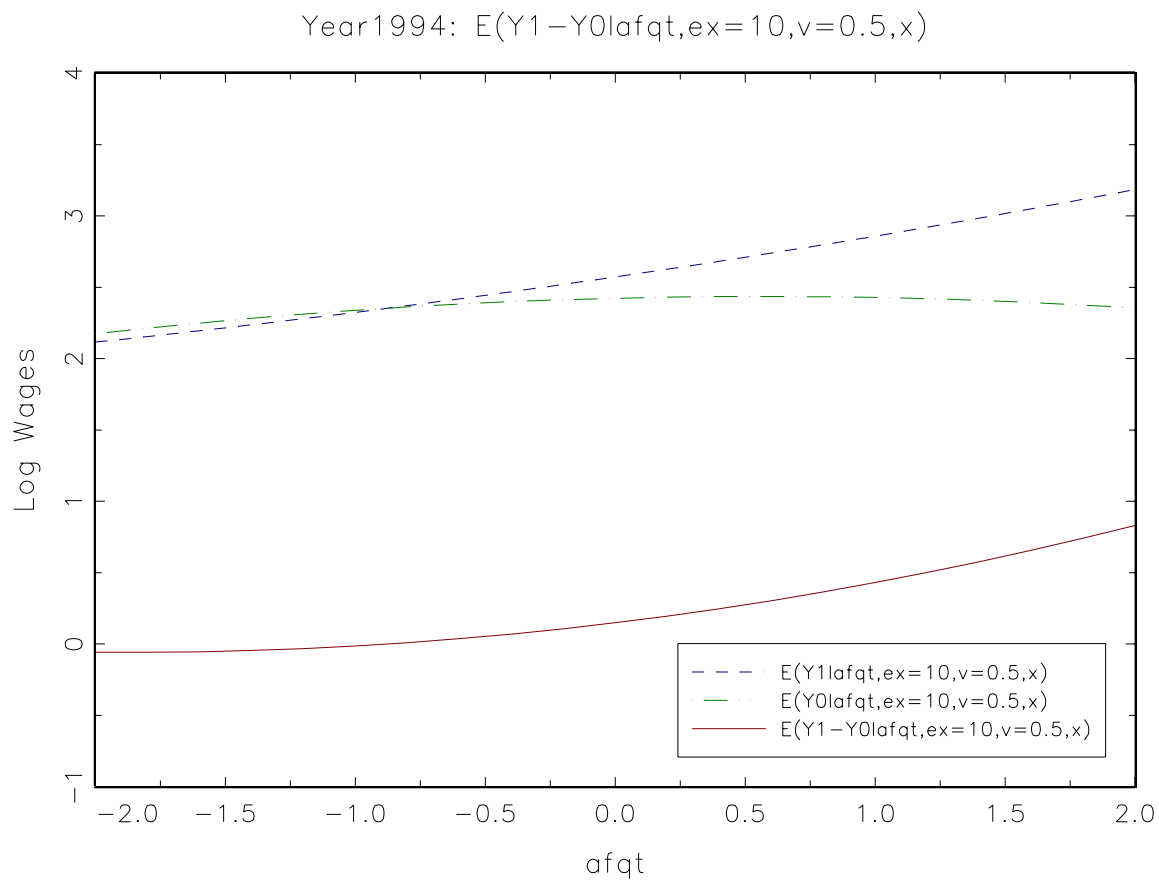


Figure 7.

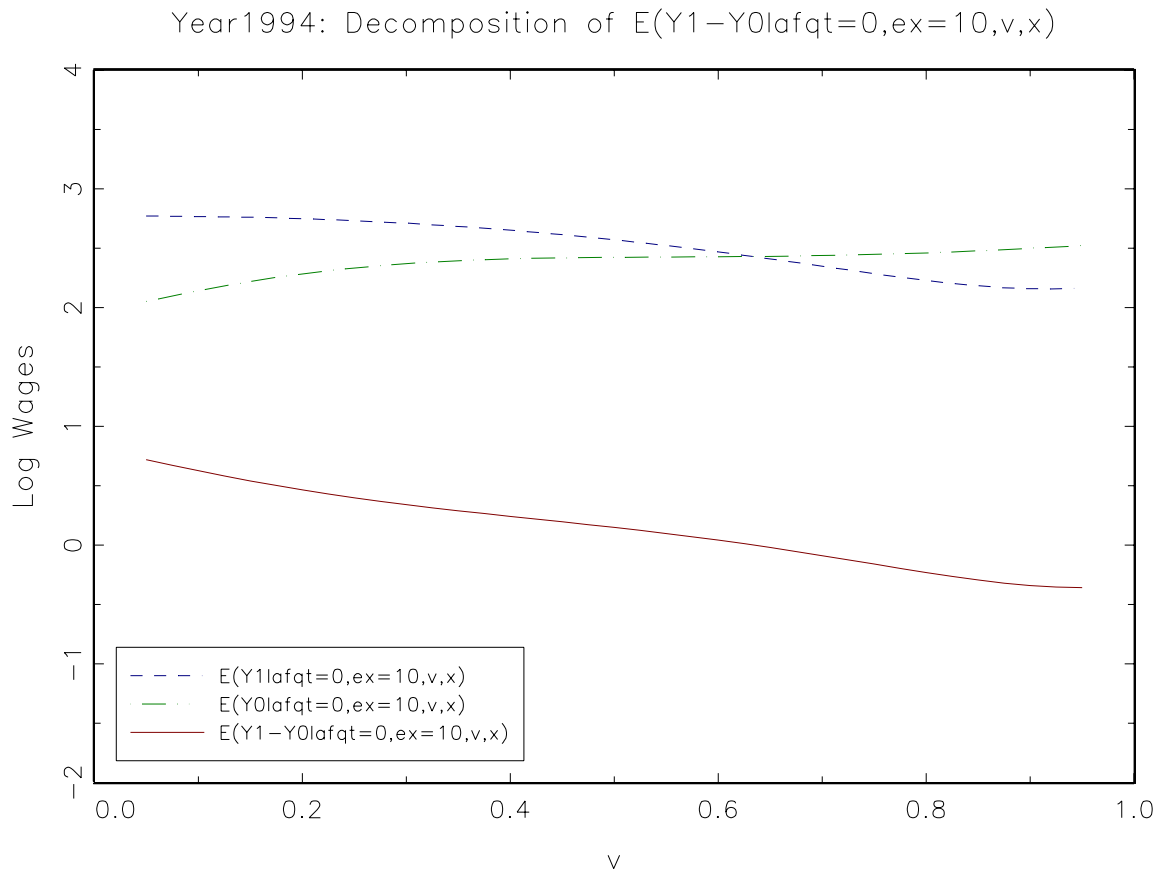


Figure 8.

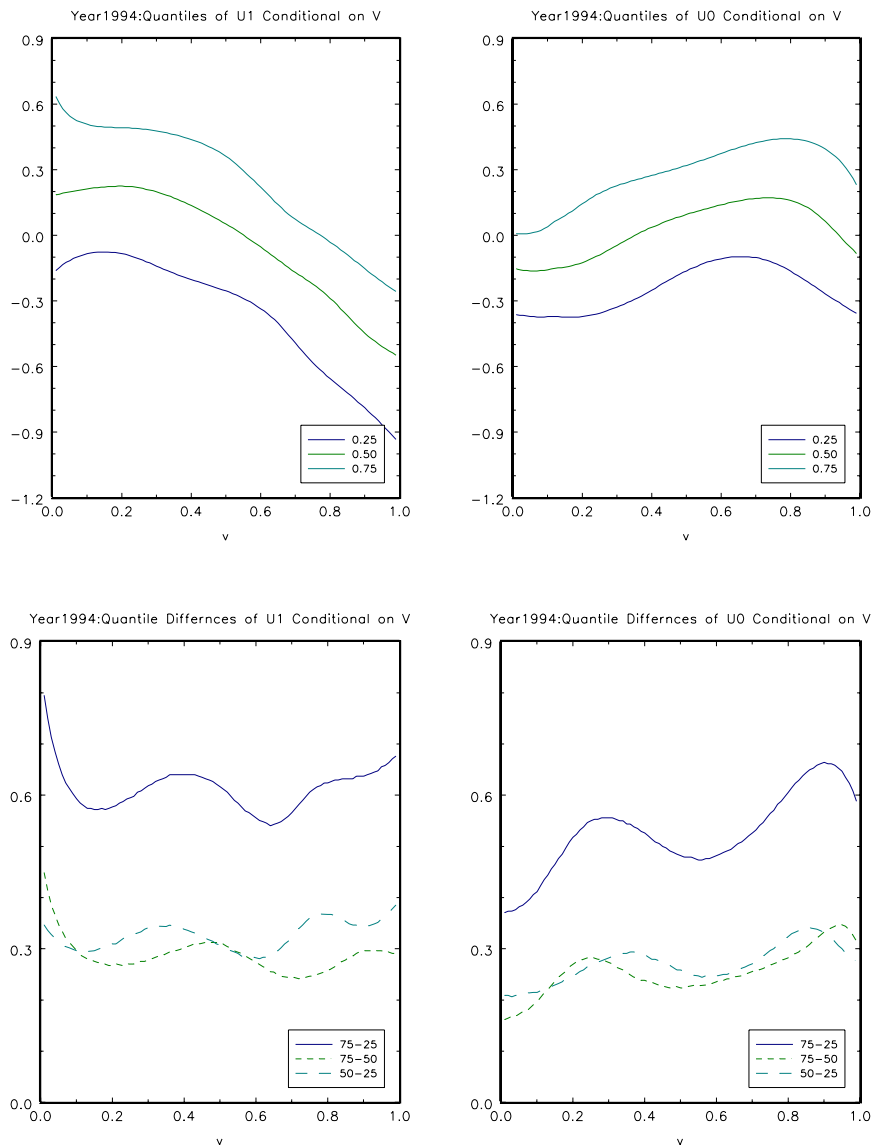


Figure 9.

Conditional Means of College and High School: Model and Data

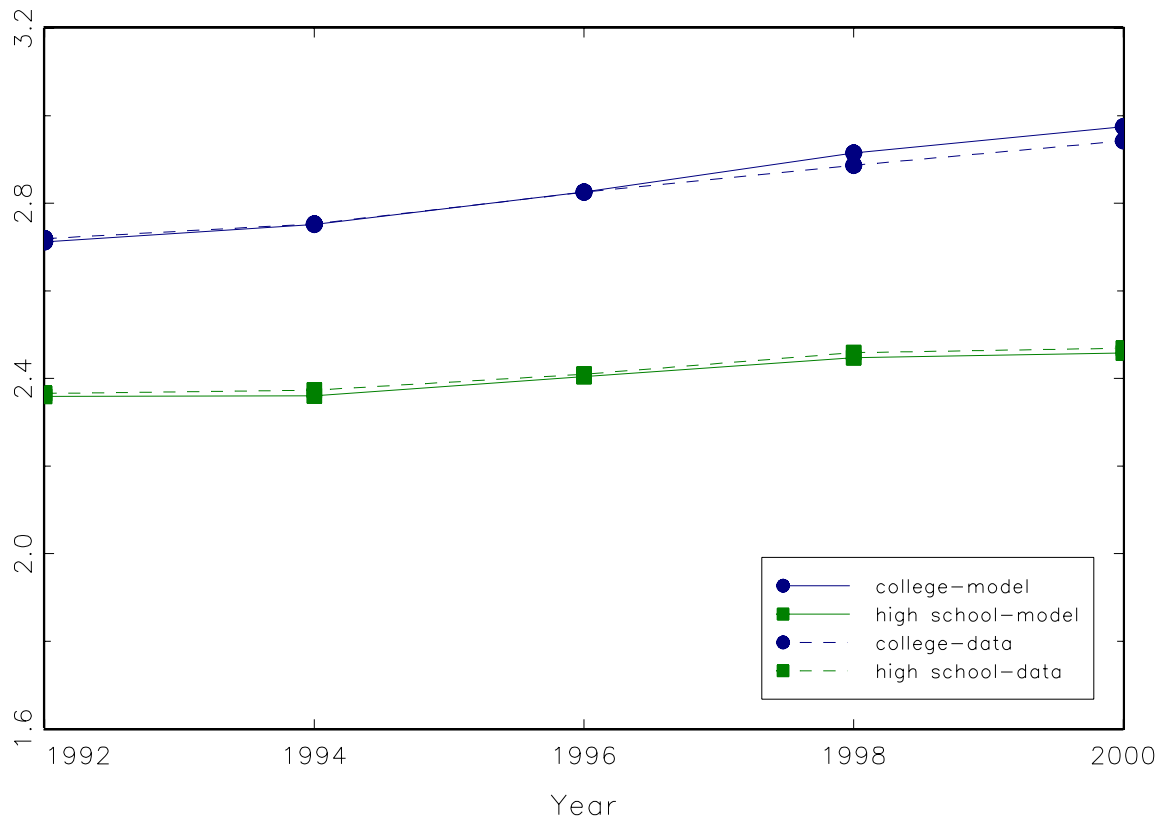


Figure 10.

Counterfactual and Conditional Means of College and High School

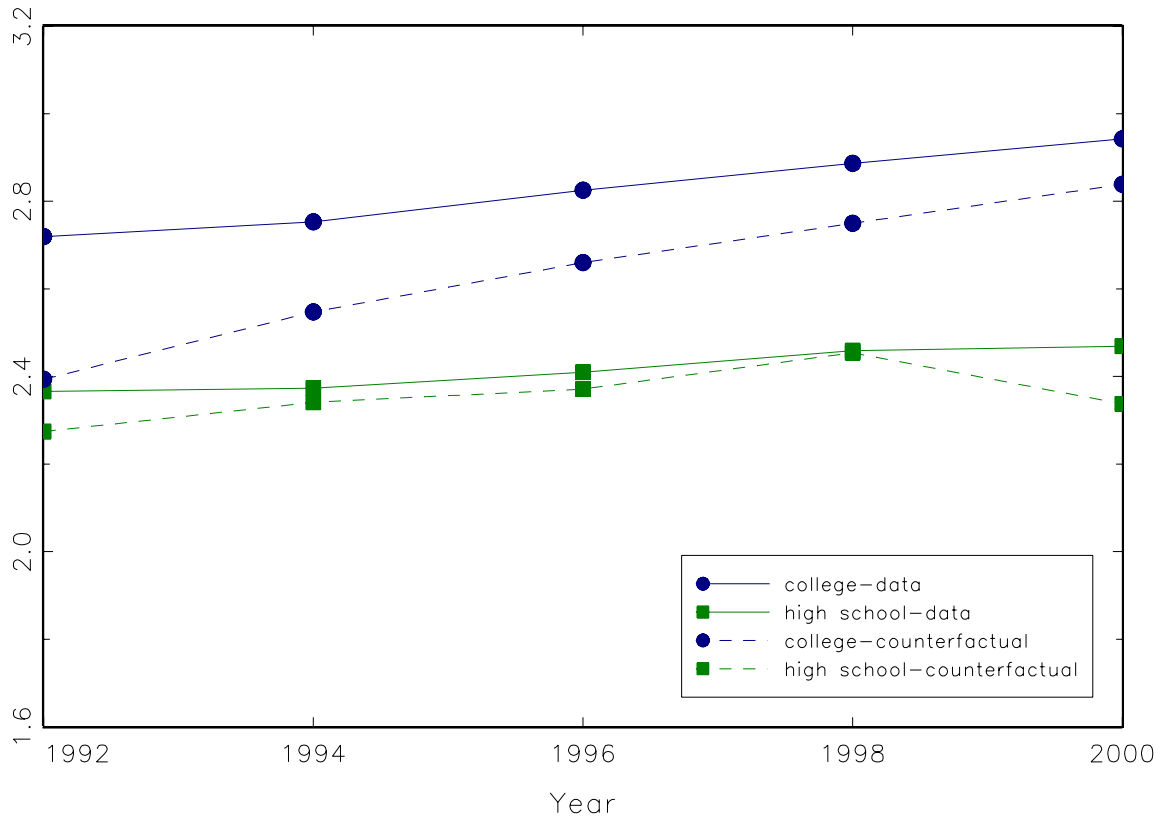


Figure 11.

Return to One Year of College: ATE, TT, TUT, and OLS

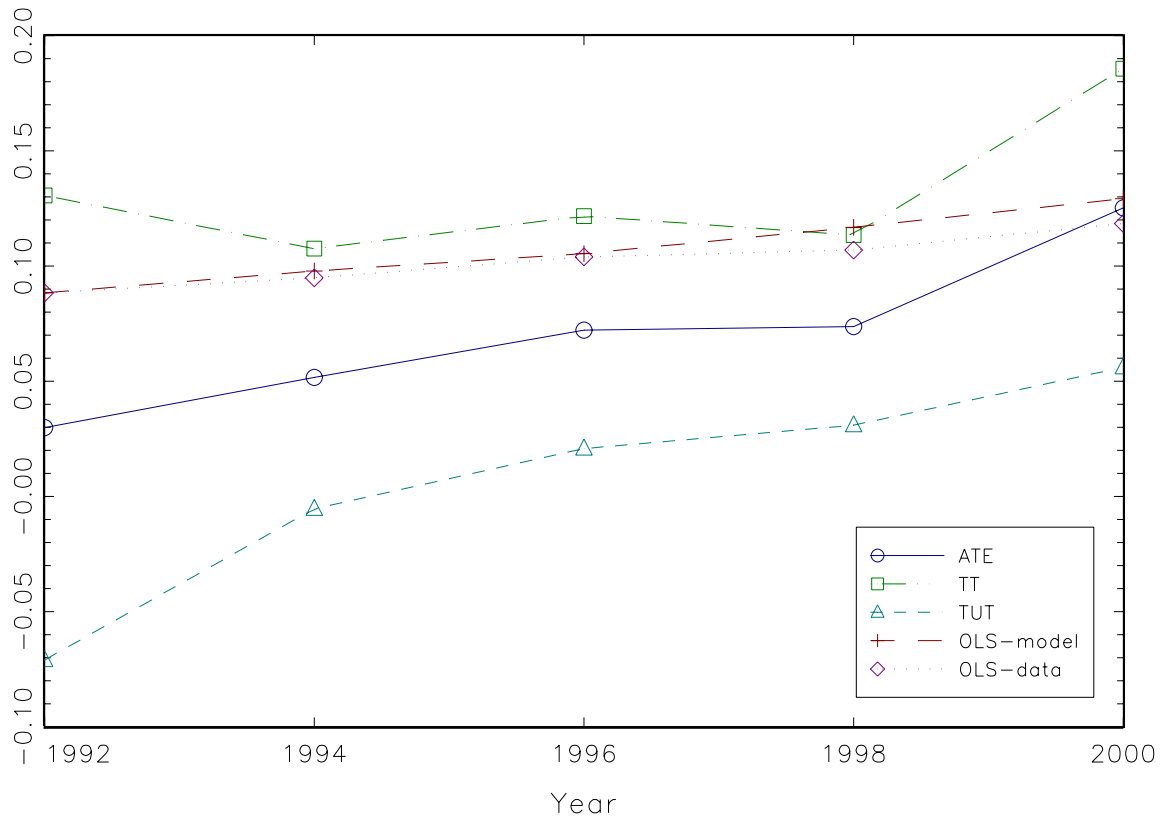


Figure 12.

Conditional Quantiles (College): Model and Data

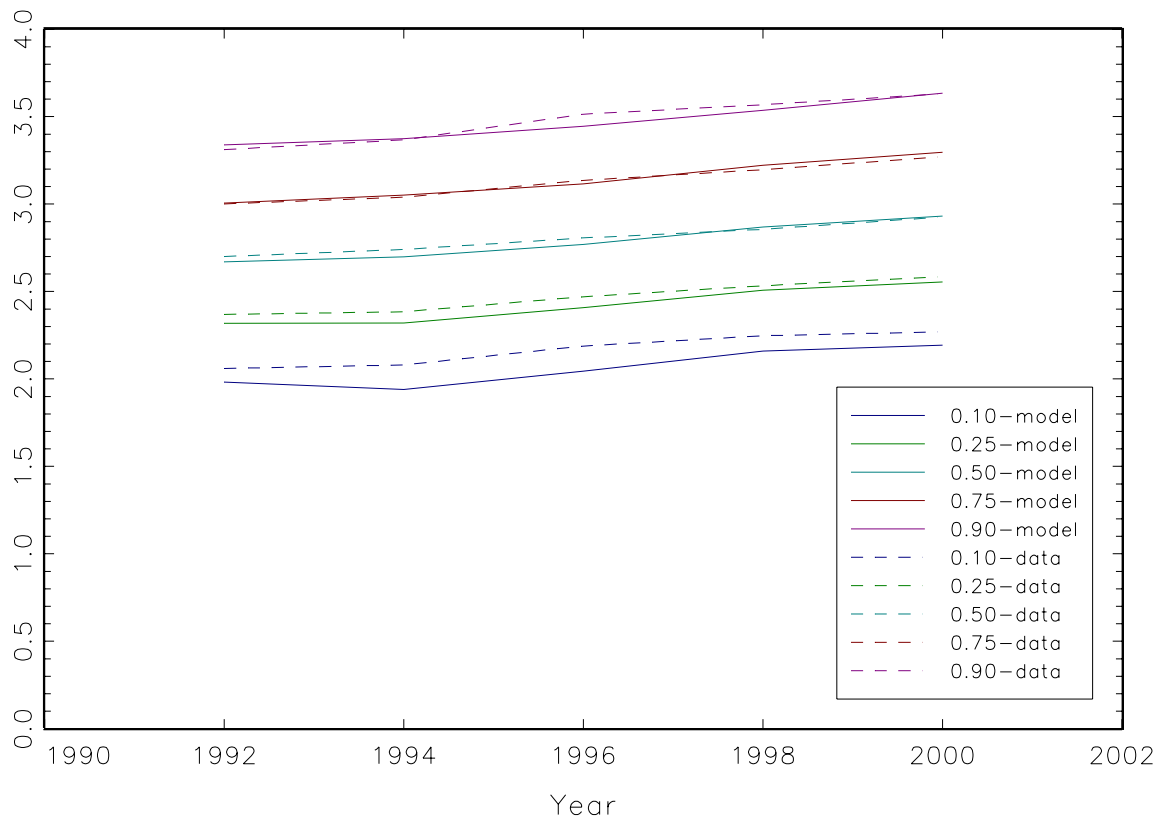


Figure 13.

Conditional Quantiles (High School): Model and Data

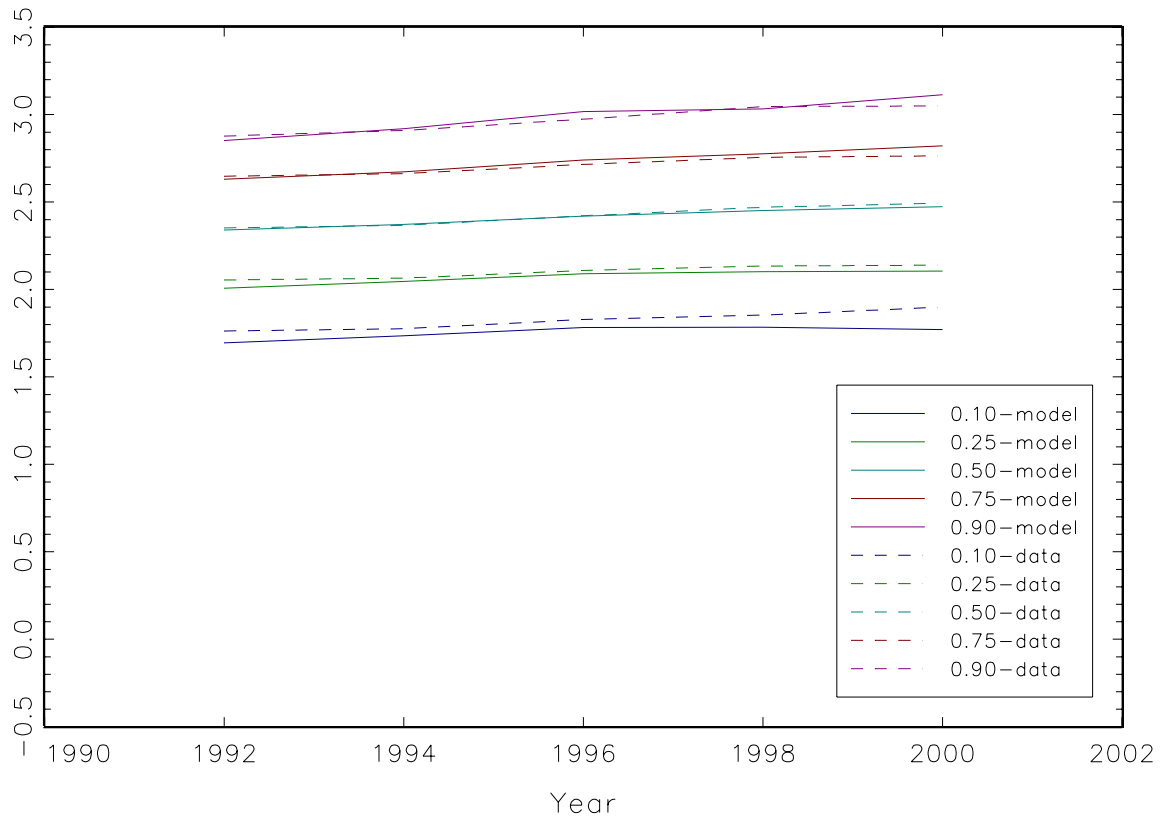


Figure 14.

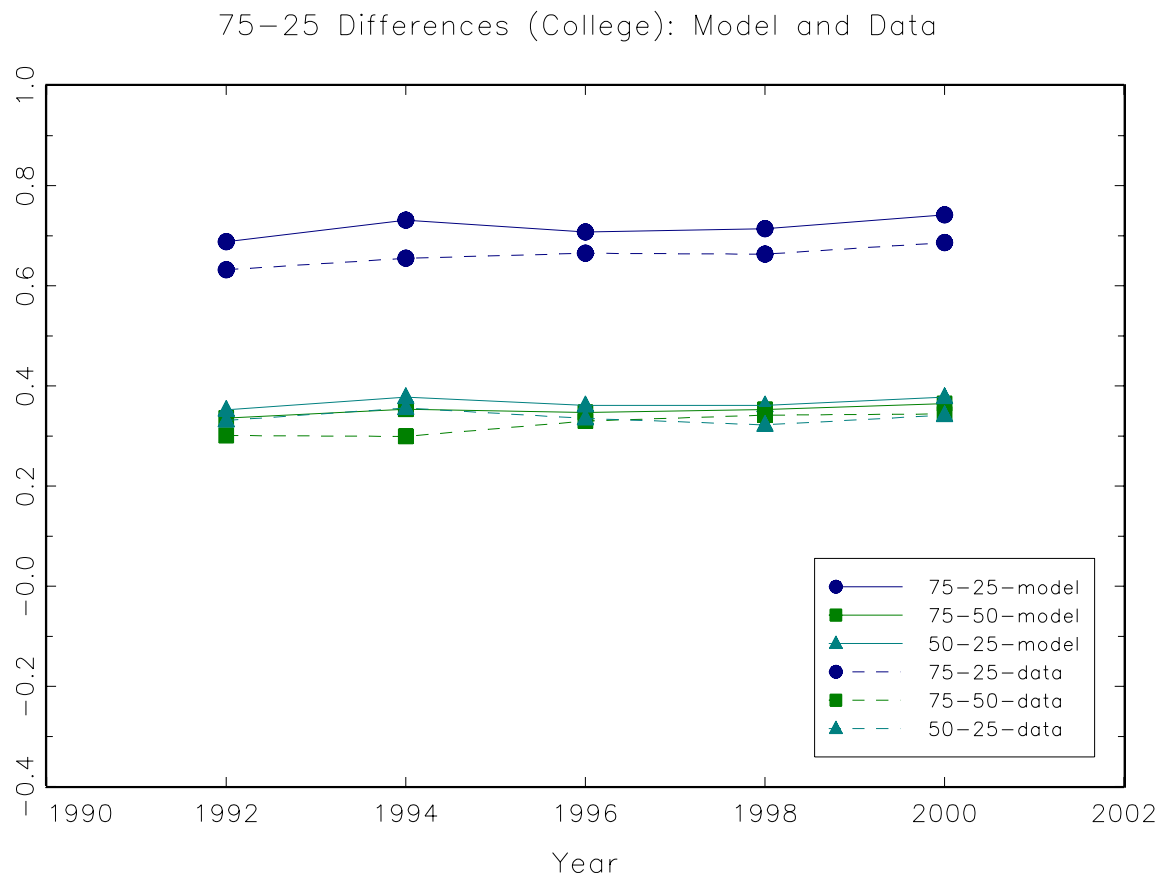


Figure 15.

75-25 Differences (High School): Model and Data

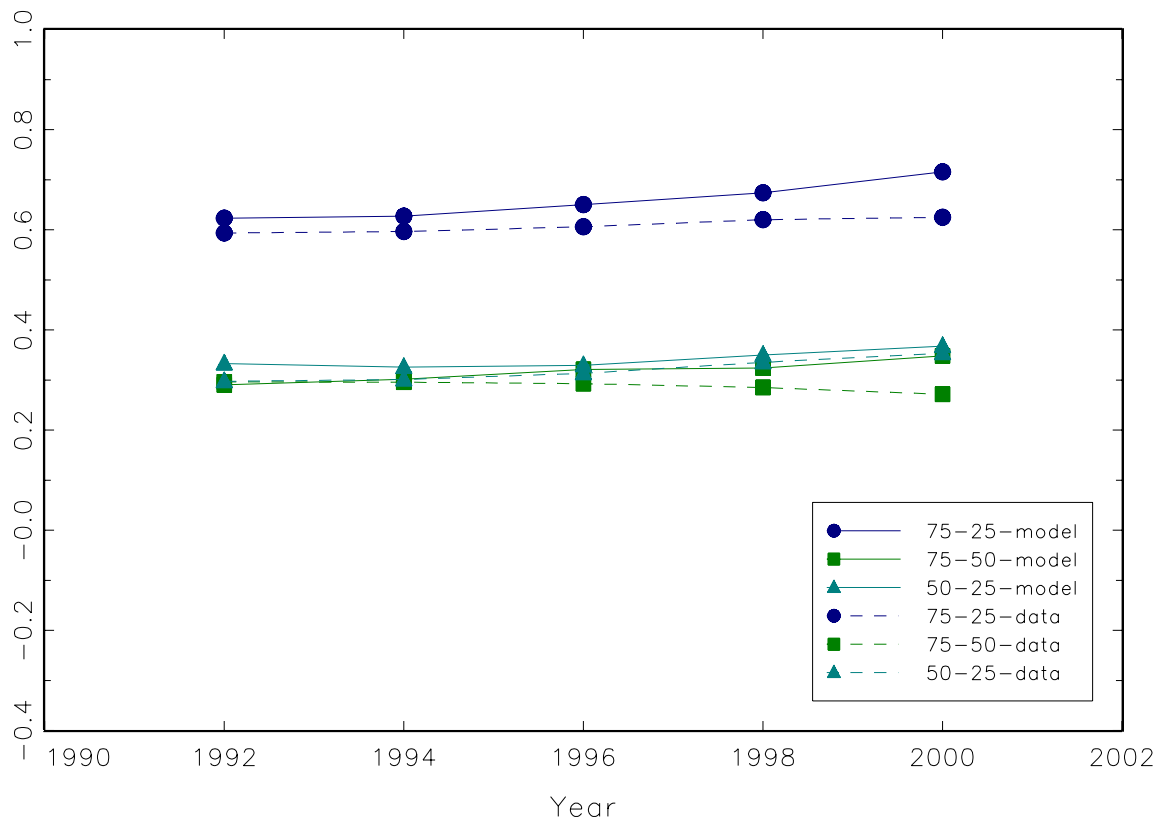


Figure 16.

75-25 Counterfactual and Conditional Differences (College)

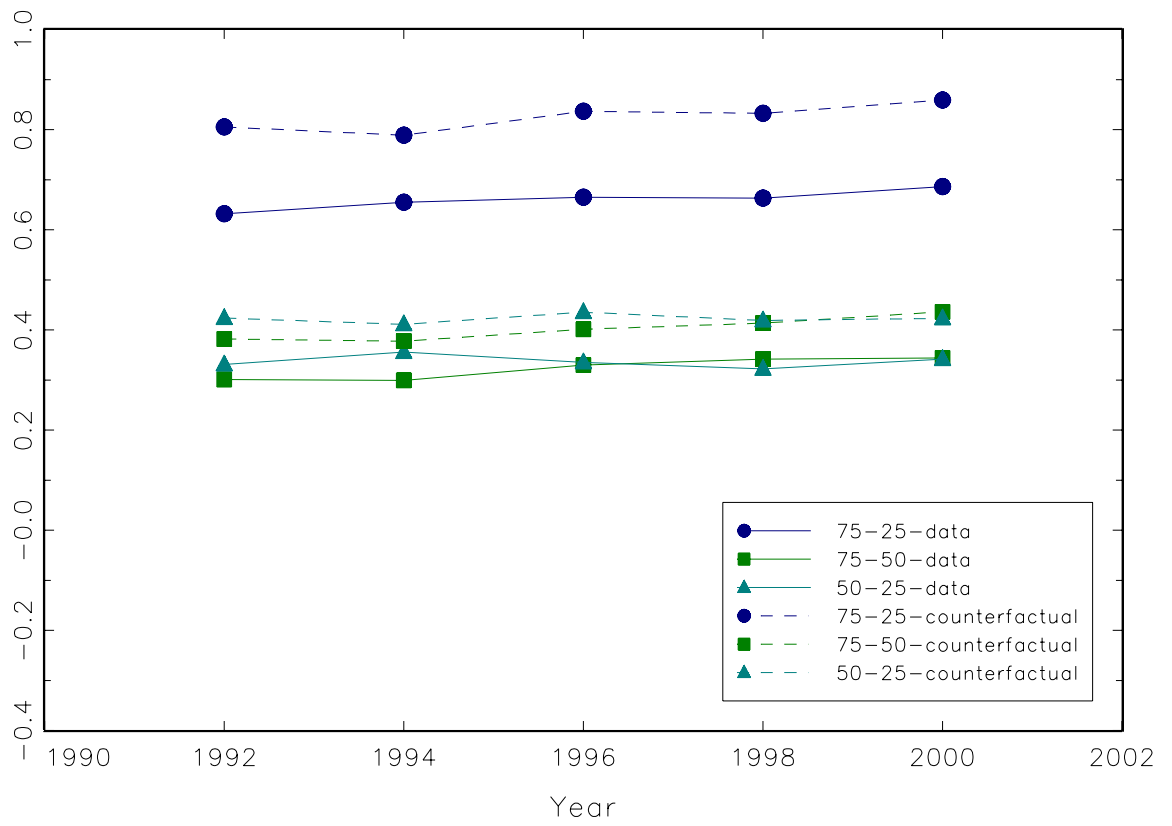


Figure 17.

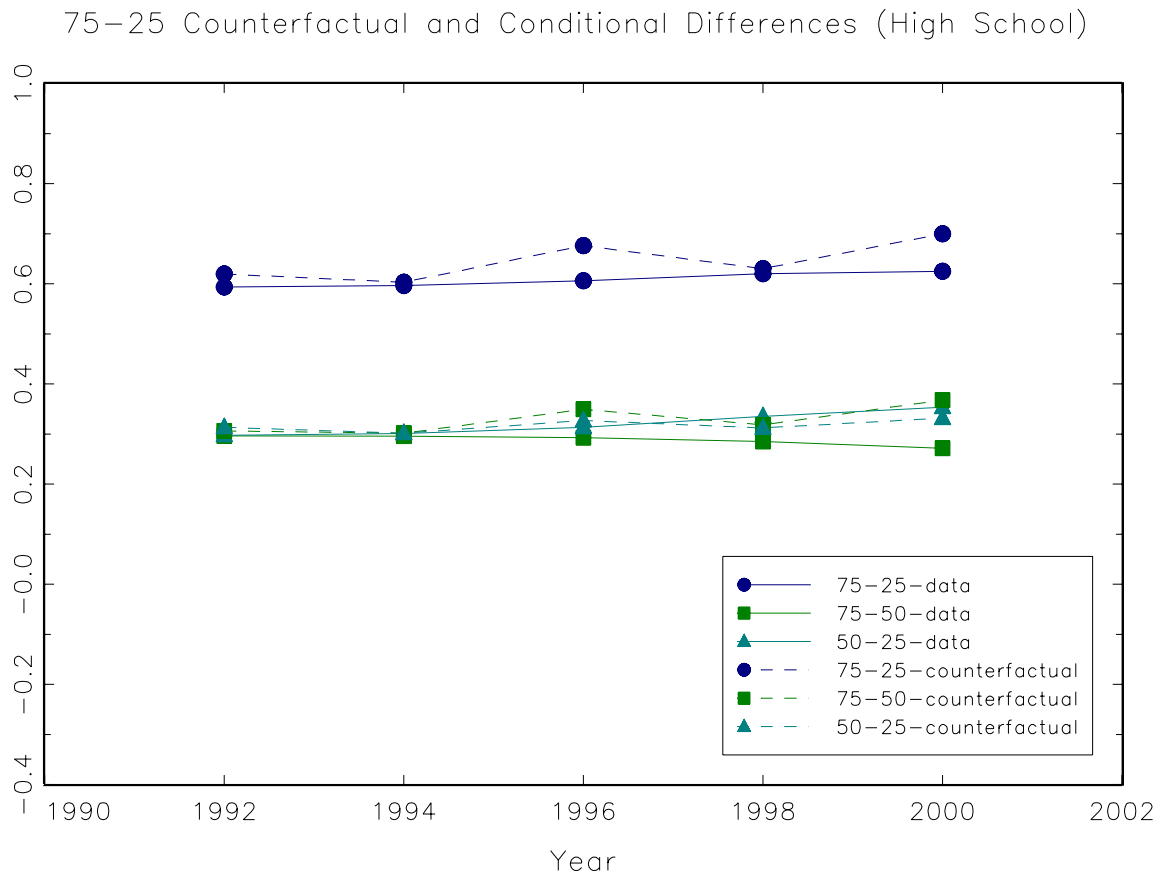


Figure 18.

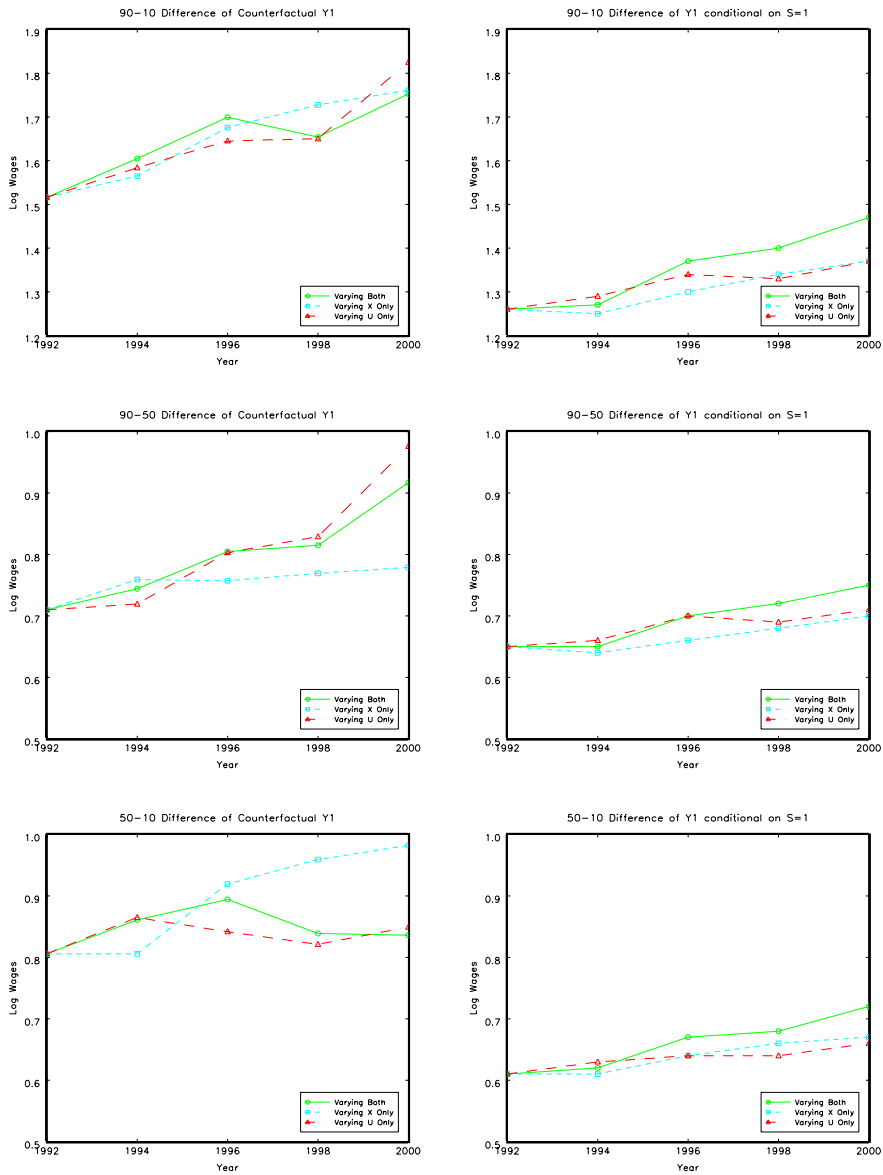


Figure 19.

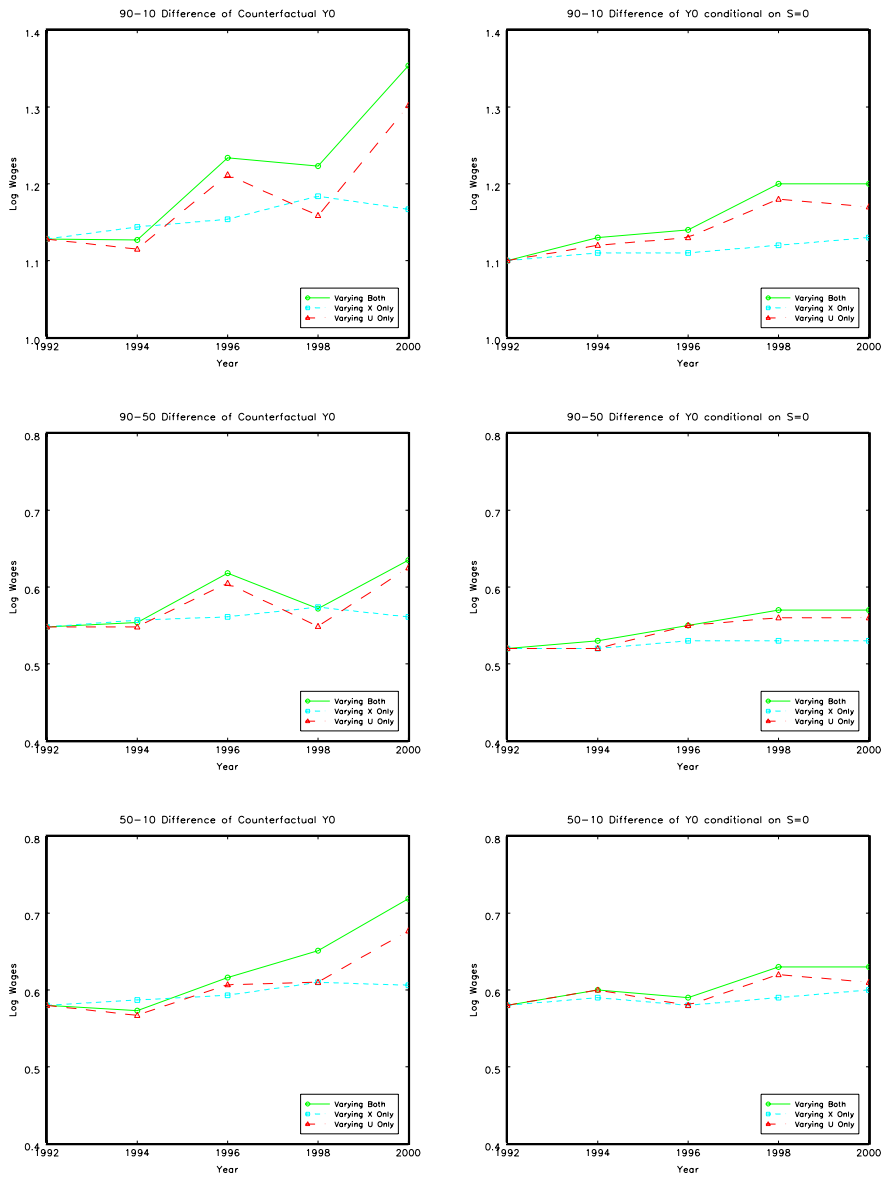


Figure 20.

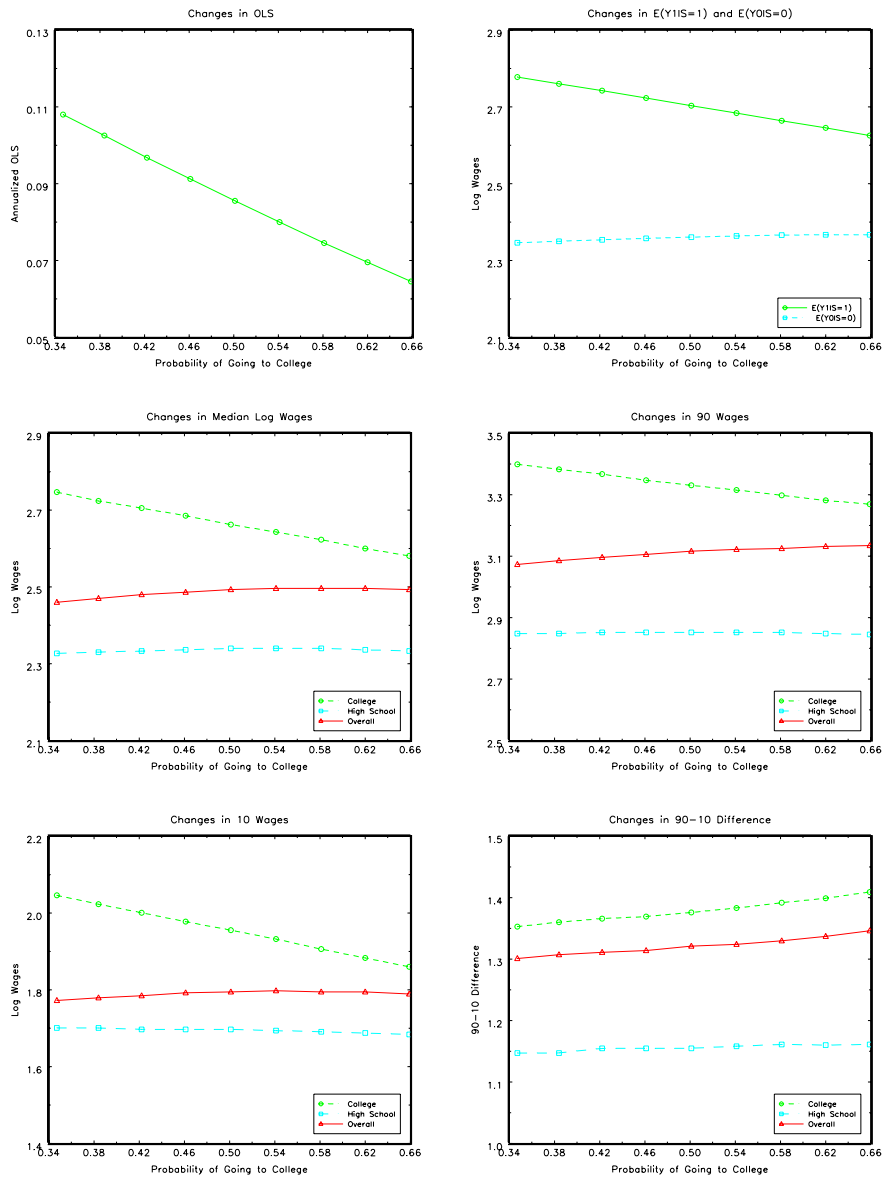


Figure 21.