Counterfactual distributions of wages via quantile regression with endogeneity

Elena Martinez-Sanchis, Juan Mora and Ilker Kandemir

Ivie

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

La Serie AD es continuadora de la labor iniciada por el Departamento de Fundamentos de Análisis Económico de la Universidad de Alicante en su colección "A DISCUSIÓN" y difunde trabajos de marcado contenido teórico. Esta serie es coordinada por Carmen Herrero.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie http://www.ivie.es, así como las instrucciones para los autores que desean publicar en nuestras series.

Counterfactual distributions of wages via quantile regression with endogeneity[*]

Elena Martinez-Sanchis, Juan Mora and Ilker Kandemir[**]

Abstract

Counterfactual decompositions allow the researcher to analyze the changes in wage distributions by discriminating between the effect of changes in the population characteristics and the effect of changes in returns to these characteristics. In this paper, counterfactual distributions are derived by recovering the conditional distribution via a set of quantile regressions, and correcting for the endogeneity of schooling decisions using a control function approach. Our proposal enables us to isolate the effect on the wage distribution of changes in both the conditional and unconditional distribution of schooling and changes in the distribution of unobserved ability. This methodology is used to analyze the sources of the changes in wage distribution that took place in the United States between 1983 and 1993, using proximity to college for different parental background as instruments. Our results show that the change in the distribution of ability had a negative effect on wages at the low quantiles, which almost compensates the positive effect of the change in the schooling distribution over this period. It is also found that the impact on wages of changes in the conditional distribution of unobserved ability is larger than the impact of changes in the conditional distribution of distance to college.

Keywords: Counterfactual Decomposition, Wage Inequality, Quantile Regression, Endogeneity.

JEL codes: C21, C14, D63, J31.

[**] E. Martinez-Sanchis, Dpto. Fundamentos del Análisis Económico, Universidad de Alicante. J. Mora, Dpto. Fundamentos del Análisis Económico, Universidad de Alicante. I. Kandemir, University College, London. Corresponding author: J. Mora, juan@merlin.fae.ua.es.

1   Introduction

Wage inequality has been an active area of interest for econometric studies since the mid-sixties, and even more so since the eighties due to the accelerating wage inequality, not only among different groups of characteristics but also among the individuals which belong to the same group. This phenomenon is well documented in Levy and Murnane (1992), Juhn et al. (1993) and Gottschalk and Smeeding (1997), among many others. The skill-biased technological change explanation to this wage inequality was based on the fact that mean regressions of wage indicated a rise in the returns to schooling despite the increase of the relative supply of high-skilled labour. However, the increase in education seems to have exacerbated both unconditional and conditional wage inequality. One possible explanation to this empirical finding is that although an increase in the educated labor supply would decrease the wage of these individuals, highly educated individuals might experience greater spreads in the wage distribution because of a different impact of schooling on wages at the top and at the bottom of the wage distribution. Other methods different from the mean regression are needed in order to disentangle the effect of changes in the composition of the labor force from the effects of changes in returns to regressors on the whole distribution of wages, and not only on the average wages.

DiNardo et al. (1996) propose a method to extend the traditional Oaxaca decomposition (Oaxaca, 1973) of the effects of changes in several explanatory factors in mean wage regressions to the entire distribution of wages. Their methodology amounts to constructing counterfactual marginal distributions in a given year, assuming that the distribution of observed individual characteristics had been the same as in the comparison year. The comparison of actual marginal distributions and counterfactual distributions allows them to asses the extent to which the observed changes in the marginal distribution of wages may be ascribed to changes in the distribution of the individual attributes.

An alternative method to estimate actual and counterfactual distributions was proposed by Machado and Mata (2005), using a series of linear quantile regressions; their procedure, which may be thought of as semiparametric, avoids the curse of dimensionality, which affects the approach suggested by DiNardo et al. (1996), and potentially reveals the whole conditional distribution of wages, if a large enough number of quantiles are estimated. Moreover, the methodology in Machado and Mata (2005) has the appealing feature that the estimated coefficients can be used to construct a measure of the residual or within-group inequality (i.e. the spread of the conditional wage distribution derived from the quantile regression model). Melly (2005) proposes a similar procedure as the one in Machado and Mata (2005), which solves the problem of the crossing of different quantile curves; additionally, he proposes a decomposition

of the changes in returns into two parts: one captures the changes in median returns, and the other captures changes due to the widening in returns across the wage distribution. Applications and extensions of the methodology of Machado and Mata (2005) include Albrecht et al. (2003), Autor et al. (2005) and Arulampalam et al. (2007), among others. Gosling et al. (2000) also propose an alternative three-step procedure to obtain unconditional quantiles from estimated conditional quantile functions.

Although controlling for endogeneity is a common problem in the estimation of wage equations in Econometrics, it has not been taken into account before in the literature on counterfactual distributions. The main objective of this paper is to extend the Machado and Mata (2005) methodology by discussing how the endogeneity problem of schooling may be solved and accounted for in the construction of counterfactual distributions. Our assumptions allow the unobservable factors affecting the education decision to be correlated with the unobserved skills affecting wages. The estimation problems that arise in this context are solved by using the control function approach as in Lee (2007). The importance of controlling for endogeneity when deriving counterfactual distributions lies in the fact that it also allows us to discuss the role that changes in the distribution of unobserved components (e.g. ability) play in the observed changes of the distribution of wages.

Our methodology likewise yields a decomposition of the part of the differences in the wage distribution which is explained by changes in the joint distribution of individual attributes; this decomposition is deeper than those proposed in the previous literature: we do not only review how to obtain counterfactual quantiles when only the distribution of one (or some, but not all) of the explanatory variables remains fixed, but we also discuss how the so-called composition effect can be further decomposed, and this helps to clarify the role of the changes in the (conditional or marginal) distribution of each covariate. The estimation of the counterfactual quantiles that we propose might involve high dimensional nonparametric estimation, which might be quite data demanding in applications with a large set of conditioning variables and samples of the usual size; for these cases, we propose parametric estimators consistent with our model for the schooling decision process. An alternative decomposition can be found in Firpo et al. (2007); they provide estimates of the effect of the change in the expectation of each regressor, when the distribution of the rest of the regressors remains fixed. In contrast, we are able to compute counterfactual effects of changes on the whole distribution of some or a particular covariate, not just on the mean.

In the empirical section, we analyse the sources of the changes in the distribution of wages in the United States between 1983 and 1993 using hourly wage data from the Panel Study of Income

Dynamics (PSID). The decomposition of the observed changes in the wage distribution between both periods into the composition effect and the effect of a change in the returns shows that the change in the joint distribution of the covariates would have shifted the wage distribution to the right if returns were held fixed at 1993. Our empirical results point out that the increase in wage inequality observed between 1983 and 1993 is chiefly explained by the change in the returns along all the quantiles, which arises due to an increase in within-group inequality. Our results suggest that the isolated effect of the change in the marginal distribution of schooling observed in the eighties would have increased all quantiles, and also inequality mainly at the bottom of the distribution; but the isolated effect of the change in the unobserved ability, conditional on education, would have been the opposite, almost compensating the positive effect of schooling at the bottom quantiles.

The rest of the paper is organized as follows. In Section 2, we introduce the underlying model, describe the estimation procedure used, and discuss the appropriateness of the stochastic assumptions underlying the control function approach. Section 3 discusses how marginal and counterfactual densities can be derived when endogeneity of schooling is taken into account, derives formally how changes in wages can be decomposed into changes in regressors and changes in returns in this context, and explains how a further decomposition may be achieved for each of these terms. Section 4 describes the data set that is used in the application. In Section 5, we apply the proposed methodology to the study of the sources of wage inequality in the United States between 1983 and 1993. Section 6 concludes.

2 Model and inference

2.1 Model

Let W be the log wage, S the years of schooling and $Z = (Z_1', Z_2')' \in \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$ the vector of regressors including individual characteristics (no constant variable is included in Z and $p_1 + p_2 \equiv p$). Hereafter, for a given $\tau \in (0,1)$, $Q_\tau(\xi)$ denotes the τth quantile of the random variable ξ. Following the control function approach proposed in Lee (2007), we assume that, for a real number τ in $(0,1)$, there exist parameters $\beta_\tau \in \mathbb{R}$, $\gamma_\tau \in \mathbb{R}^{p_1}$, $\mu \in \mathbb{R}$ and $\pi \in \mathbb{R}^p$ such that

$$\left. \begin{array}{l} W = \beta_\tau S + \gamma_\tau' Z_1 + U_\tau, \\ S = \mu + \pi' Z + V, \end{array} \right\} \tag{2.1}$$

where U_τ and V are unobserved random variables such that, for any $(z', v)'$ in the support of $(Z', V)'$,

$$Q_\tau(U_\tau \mid Z = z, V = v) = Q_\tau(U_\tau \mid V = v) \qquad \text{and} \tag{2.2}$$

$$Q_{0.5}(V \mid Z = z) = 0. \tag{2.3}$$

Note that U_τ and V are not assumed to be independent and, hence, schooling is an endogenous variable. It is easy to prove that assumption (2.2), together with the second equation in (2.1), imply that $Q_\tau(U_\tau \mid S = s, Z_1 = z_1, V = v) = Q_\tau(U_\tau \mid V = v)$; hence, if we denote $\lambda_\tau(v) \equiv Q_\tau(U_\tau \mid V = v)$, it follows from (2.1)-(2.3) that

$$Q_\tau(W \mid S = s, Z_1 = z_1, V = v) = \beta_\tau s + \gamma_\tau' z_1 + \lambda_\tau(v) \qquad \text{and} \tag{2.4}$$

$$Q_{0.5}(S \mid Z = z) = \mu + \pi' z. \tag{2.5}$$

Note that (2.4) provides an explicit expression for the conditional quantile of wages as a function of schooling, the control variables affecting wage and the unobservable component affecting schooling decisions. The control variables in Z_1 may include age, tenure and powers of age and/or tenure among other demographic variables. The variables included in Z_2 ("instruments") are excluded variables so that they affect the schooling decisions but not the individual wages, such as cost shifters involved in the educational decisions; in Section 4 we discuss our choice of instruments. Note that, following Lee (2003), identification for β_τ, γ_τ and $\lambda_\tau(\cdot)$ is achieved if $\mathrm{Var}(S, Z_1 \mid V)$ is nonsingular almost surely (rank condition); this condition would not hold if the last p_2 components of π, i.e. those associated to Z_2, are all equal to 0 (significance of instruments).

Although we interpret throughout the paper that the unobservable random variable V captures individual ability, this is obviously just a simplification, as many other unobservables can affect schooling after controlling for Z, such as preferences. The unobservables in the wage equation U_τ are also interpreted as luck in the labour market, though many other sources of heterogeneity can affect wages given both schooling and the demographic characteristics Z_1. Condition (2.3) implies that unobserved ability is median independent of Z. Condition (2.2) assumes that the unobservables affecting wages are τth-quantile independent of the exogenous variables Z, once we control for V.

Schooling might be interpreted as an equilibrium outcome partially determined by U_τ and, since U_τ and V have not been assumed to be independent, then years of education is an endogenous variable. In order to motivate the particular form of the system of equations above, consider that the agent optimally chooses S by maximizing the expected outcome minus the cost associated to each particular level of schooling, given by the information set of the individual. We assume this information set contains only ability or preferences captured by V, cost shifters included in Z_2 and demographic variables included in Z_1. Let $\phi(S, Z_1, U_\tau)$ be the wage function linearly specified in the first equation in (2.1) and $c(S, Z)$ be the cost function associated to

the level of schooling S. Then S is obtained as the solution to an individual choice problem: $S = \arg\max_s \mathbb{E}\{\phi(s, Z_1, U_\tau) \mid V, Z\} - c(s, Z)$. This motivates that the schooling equation in (2.1) is a reduced form which is a function of Z and unobserved V.

2.2 Discussion of the stochastic assumptions

Endogeneity of schooling could have arisen under alternative stochastic assumptions on the unobservables. More specifically, the quantile restriction in (2.2), which characterizes the so-called control function (CF) approach, could be replaced by the quantile restriction $Q_\tau(U_\tau \mid Z = z) = Q_\tau(U_\tau)$, which characterizes the so-called instrumental variable (IV) approach (see e.g. Honoré and Hu, 2004). The IV assumption means that the unobservables affecting wages are τth quantile independent of the variables in Z. Schooling in this case is endogenous since S does not appear in the set of conditioning variables. As opposed to the CF assumption, there is no need to condition on ability to assume quantile independence between luck in the labour market and demographics and cost shifters in Z. Thus, the IV assumptions mean that those individuals for which higher education is less costly are not more or less lucky in the labour market than those individuals for which costs are higher. Under the CF restrictions, this quantile independence only holds once we condition on a particular ability level. In fact, the IV assumption does not provide information about the dependence of the unobservables in the wage and the schooling equation. Only in linear conditional mean models where the endogenous variable appears linearly in the equation and a linear reduced form is estimated for the endogenous variable, the CF approach imposes extra assumptions not imposed by the IV approaches which lead to less robust estimators derived under CF assumptions. However, for more general structural models, the CF restriction (2.2) and this IV restriction are non-nested and both sets of restrictions are no more or less general. As stated by Blundell and Powell (2003) and Lee (2004), both sets of assumptions are implied if the strong independence of (U_τ, V) and Z is satisfied.

Also note that the CF approach and the IV approach allow for different types of heteroskedasticity: the sufficient conditions that a data generation process must satisfy in order to ensure that the CF restriction holds do not exclude the possibility of heteroskedasticity depending on V, S and Z_1; and the sufficient conditions ensuring that the IV restriction is satisfied only allow for heteroskedasticity depending on Z_1 (Lee, 2004). In our context, given the sequential nature of the relationship between schooling and wages, it seems more reasonable to assume that the heterogeneity of the unobservables in the wage equation depends on those unobservables affecting schooling, on schooling itself and on the exogenous variables Z_1.

Chernozhukov and Hansen (2005, 2006) show that the IV restriction, together with the assumption that the unobserved ability variable is made invariant to schooling (conditional on Z_1), identifies β_τ and γ_τ. However, only the CF restriction allows us to identify the conditional quantile $Q_\tau(W \mid S = s, Z_1 = z_1, V = v)$, and this will allow us to assess the impact of a change in the distribution of, say, the unobservables affecting schooling V, while keeping fixed schooling and the control variables Z_1. In this sense, CF restriction allows to solve for the endogeneity of schooling as an omitted variable problem by including an unknown function of the unobservable ability. This kind of counterfactual analysis is not possible under the IV restrictions. Nonetheless, the IV assumptions imply the advantage that there is no need for specifying a reduced form equation for S and no continuity assumptions on S are required.

Our model does not cover the case of heterogenous returns to schooling depending on either the unobserved ability V and/or the unobserved U_τ. Under the CF assumption, when the returns to schooling are heterogenous in both V and U_τ, the estimation of the returns to schooling at different values of the support of ability could be achieved in our model by adding interactions of an unknown function of V with schooling, for each quantile τ (see Lee 2007). Under the strong assumptions of independence of U_τ and V, and allowing for heterogeneity in both ability and luck, the quantile treatment effect of education on earnings at different ability quantiles and luck quantiles has been considered by Brunello et al. (2009) (see also Ma and Koenker 2006 and Chesher 2003 for the estimation).

Our model assumes that the conditional median of S given $Z = z$ is linear (see 2.5). Instead, we could have assumed that the αth conditional quantile of S given $Z = z$ is linear for some other quantile $\alpha \in (0, 1)$. In this case, in practice one would have to decide which α should consider. Since our main concern is the estimation of the conditional quantile of wages, one could try to find the value α which leads to the most efficient estimation of this conditional quantile. If Z and $V_\alpha \equiv S - Q_\alpha(S \mid Z)$ were independent, this criterion would lead to choose the value $\alpha \in (0, 1)$ that minimizes $\alpha(1 - \alpha)/f_{V_\alpha}(0)^2$, where $f_{V_\alpha}(\cdot)$ denotes the probability density function of V_α (see Lee 2007). However, in our context it is likely that Z and V_α are not independent (e.g. this hypothesis is rejected in our empirical application, see Section 5), and in this case the previous criterion leads to an intractable optimization problem. Hence, for clarity of exposition, we have decided to state our assumptions considering the conditional median for the schooling equation.

2.3 Estimation

Given independent and identically distributed (i.i.d.) observations $\{(W_i, S_i, Z_i)\}_{i=1}^N$, Lee (2007) describes how the parameters of the model and the unknown function $\lambda_\tau(\cdot)$ can be consistently

estimated. We briefly describe this estimation procedure with the notation that is more appropriate for the discussion of the following section.

First, the conditional quantile restriction in (2.5) suggests to estimate μ and π using linear median regression estimation; thus, the first stage estimates, denoted by $\widehat{\mu}$ and $\widehat{\pi}$, are the solution to the optimization problem

$$\min_{(\mu,\pi)} \sum_{i=1}^{N} \left| S_i - \mu - \pi' Z_i \right|, \tag{2.6}$$

In a second stage, the conditional quantile restriction in (2.4) suggests that, once μ and π are estimated, β_τ and γ_τ can be estimated using a series approximation of $\lambda_\tau(\cdot)$ and a trimmed linear quantile regression; this trimming is introduced not only to show the uniform convergence and asymptotic normality of the estimator, but also to avoid influences of large values of the covariates (Newey et al. 1999). Thus, the second stage estimates, denoted by $\widehat{\beta}_\tau$ and $\widehat{\gamma}_\tau$, are the first $1 + p_1$ components of the solution to the optimization problem

$$\min_{(\beta,\gamma,\delta)} \sum_{i=1}^{N} \left| W_i - \beta S_i - \gamma' Z_{1i} - \delta' P(S_i - \widehat{\mu} - \widehat{\pi}' Z_i) \right| t(S_i, Z_i, \widehat{\mu}, \widehat{\pi}), \tag{2.7}$$

where $P(\cdot) \equiv (p_1(\cdot), ..., p_{k_N}(\cdot))'$, $\{p_k(\cdot)\}_{k=1}^{\infty}$ is a basis for smooth functions such that a linear combination of them can approximate $\lambda_\tau(\cdot)$, $k_N = O(N^\nu)$ for some $\nu > 0$, $t(s, z, \mu, \pi) \equiv 1\{s \in \Delta_S\}1\{z_1 \in \Delta_{Z_1}\}1\{s - \mu - \pi'z \in \Delta_V\}$, where $1\{\cdot\}$ is the usual indicator function, Δ_S and Δ_V are compact intervals in $\mathbb{R}$, and Δ_{Z_1} is a compact rectangle in $\mathbb{R}^{p_1}$. The second stage also yields an estimate of the function $\lambda_\tau(\cdot)$, namely $\widehat{\delta}'_\tau P(S_i - \widehat{\mu} - \widehat{\pi}' Z_i)$, where $\widehat{\delta}_\tau$ denotes the last k_N components of the solution to the optimization problem (2.7); however it is difficult to perform inference with this estimate. As an alternative, since $\lambda_\tau(v)$ is the τth quantile of the conditional distribution $U_\tau \mid V = v$, it can be estimated with a local linear quantile regression estimation of $\widehat{U}_\tau \equiv W - \widehat{\beta}_\tau S - \widehat{\gamma}'_\tau Z_1$ on $\widehat{V} \equiv S - \widehat{\mu} - \widehat{\pi}' Z$; thus, for any v in the support of V, the third stage estimate of $\lambda_\tau(v)$, denoted by $\widehat{\lambda}_\tau(v)$, is the first component of the solution to the optimization problem

$$\min_{(d_1,d_2)} \sum_{i=1}^{N} \left| W_i - \widehat{\beta}_\tau S_i - \widehat{\gamma}'_\tau Z_{1i} - d_1 - d_2(S_i - \widehat{\mu} - \widehat{\pi}' Z_i - v) \right| K_h(S_i - \widehat{\mu} - \widehat{\pi}' Z_i - v), \tag{2.8}$$

where $K_h(\cdot) \equiv h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function and h is a bandwidth (Yu and Jones, 1993). After this three-stage estimation procedure, it follows from (2.4) that the natural estimate of the conditional quantile function $Q_\tau(W \mid S = s, Z_1 = z_1, V = v)$ is

$$\widehat{Q}_\tau(W \mid S = s, Z_1 = z_1, V = v) = \widehat{\beta}_\tau s + \widehat{\gamma}'_\tau z_1 + \widehat{\lambda}_\tau(v). \tag{2.9}$$

This estimator plays a key role in the procedure to construct counterfactual quantiles that is described in Section 3.

The asymptotic properties of the second-stage estimators $\widehat{\beta}_\tau$ and $\widehat{\gamma}_\tau$ are derived in Lee (2007) under relatively standard smoothness and moment conditions; finally, since $\widehat{\mu}$, $\widehat{\pi}$, $\widehat{\beta}_\tau$ and $\widehat{\gamma}_\tau$ are root-N-consistent, the asymptotic properties of the third-stage estimator $\widehat{\lambda}_\tau(v)$ are the same as those of the infeasible estimator of $\lambda_\tau(v)$ that would be obtained using μ, π, β_τ and γ_τ instead of their estimates in (2.8), and the properties of this infeasible estimator are also well-known (see e.g. Chaudhuri, 1991).

In practice, the second stage requires the choice of the family of functions $\{p_k(\cdot)\}_{k=1}^\infty$, the number k_N of functions $p_k(\cdot)$ that are used in the optimization problem (2.7), and the sets Δ_S, Δ_{Z_1}, Δ_V that characterize the trimming function. The asymptotic results derived in Lee (2007) assume that $\{p_k(\cdot)\}_{k=1}^\infty$ are either polynomials or splines; the simulation results reported in Lee (2003) indicate that k_N values between 3 and 8 can be adequate, and suggest that the performance of the estimator is not very sensitive to this choice. As regards the sets Δ_S, Δ_{Z_1}, Δ_V, they must be contained in the interior of the support of S, Z_1, V, respectively; in practice they should be chosen in such a way that there are no unduly influential observations.

It is worth discussing an additional issue of practical relevance that might arise in the first stage estimation. Observe that the dependent variable in the first stage, i.e. schooling S, is continuous in nature, but possibly recorded as discrete. In almost all samples, it only takes on integer values and there are many observations with the same recorded years of education. The fact that some observations of a continuous variable take on the same value poses a problem in many contexts; but in this case this is especially worrying, since the standard optimization procedures for quantile regression assume a continuous dependent variable and do not work properly when the number of observations sharing the same value of the dependent variable is relatively high. Furthermore, the asymptotic results derived in Lee (2007) require that this first-stage endogenous variable be continuously distributed. To circumvent this problem, we suggest to transform the integer-valued variable years of schooling into a continuous variable by adding a uniform $(-0.5, 0.5)$ random noise to the recorded values. This approach is basically the same that is proposed in Machado and Santos-Silva (2005), who studied both theoretically and by simulations the consequences of this type of smoothing. The good asymptotic behaviour of the resulting estimates only holds if the conditional quantile of the smoothed dependent variable is linear on the regressors; however, the asymptotic results that they report suggest that, even if this condition is not met, the resulting estimates are surprisingly accurate and compare very favorably with respect to the estimates that are obtained if the problem is solved with other smoothing techniques. In the empirical analysis of this paper, the schooling variable is smoothed by adding a uniform $(-0.5, 0.5)$ noise, but very similar results are obtained if this variable is

smoothed by adding a uniform $(-\varepsilon, \varepsilon)$ noise, if $0.3 < \varepsilon < 1$.

3 Counterfactual quantiles with endogenous regressors

Hereafter we assume that we are interested in analysing changes in the wage distribution along T periods (or groups) when a random sample of size N_r is available for each $r \in \{1, ..., T\}$. Furthermore, we assume that in all periods and for every $\tau \in (0,1)$, assumptions (2.1), (2.2) and (2.3) hold. In the remainder of this section the set of explanatory variables $(S, Z_1', V)'$ is denoted as X.

Given a set of explanatory variables X, the starting point of DiNardo et al. (1996) to derive counterfactual densities of wages is the relationship $f_W(w) = \int f_{W|X=x}(w) \mathrm{d}F_X(x)$, where, hereafter, $F_\zeta(\cdot)$ and $f_\zeta(\cdot)$ denote the cumulative distribution function (CDF) and probability density function (PDF) of the random variable ζ, respectively. Unfortunately, when working with quantiles this type of relationship does not hold, i.e. the τth marginal quantile marginal of wages $Q_\tau(W)$ does not coincide with $\int Q_\tau(W \mid X = x) \mathrm{d}F_X(x)$. But it is still possible to derive a relationship between $Q_\tau(W)$, the marginal distribution of X and the quantiles of $W \mid X = x$; more specifically, it is easy to prove that $Q_\tau(W)$ can be defined implicitly as the solution q to the equation

$$\int \left(\int 1\{w \le q\} f_{W|X=x}(w) \mathrm{d}w \right) \mathrm{d}F_X(x) = \tau. \tag{3.1}$$

In the next subsections we describe how to use this relationship to define counterfactual quantiles.

Equation (3.1) could also be used as the starting point to derive estimates of the counterfactual quantiles (see Melly 2005). However, following Machado and Mata (2005), in this context it is easier to apply a simulation-based procedure. Here we describe how this procedure is used to derive estimates of the marginal quantiles of wages that are consistent with the model; in the next subsections we describe how it can be used to estimate counterfactual quantiles. The key idea behind the Machado and Mata (2005) procedure is the inverse probability integral transform. More specifically, for a fixed x, the inverse probability integral transform ensures that if ξ is a random variable with uniform distribution on $(0,1)$, then the random variable $Q_\xi(W \mid X = x)$ has the same distribution as $W \mid X = x$; using this result it is possible to prove that the CDF of the random variable $Q_\xi(W \mid X)$ is the same as the CDF of W. As a consequence, given our model for all conditional quantiles $Q_v(W \mid X = x)$ and their consistent estimates $\widehat{Q}_v(W \mid X = x)$ defined in (2.9), the following simulation-based procedure can be used to derive an estimate of $Q_\tau(W)$, hereafter denoted as $\widehat{Q}_\tau(W)$: i) obtain a draw of size m from the uniform distribution on $(0,1)$, denoted as $\{\xi_i\}_{i=1}^m$; ii) obtain a draw of size m from the empirical distribution $\widehat{F}_{\widehat{X}}(\cdot)$, denoted as $\{\widehat{X}_{*i}\}_{i=1}^m$ (note that this is just a draw of size m with

replacement from $\{\widehat{X}_i\}_{i=1}^N$; also note that $\widehat{X}$ is used instead of X since one of its components (e.g. V) is not observable and the correponding components should be replaced by the appropriate estimates); iii) find the τth sample quantile of $\{\widehat{Q}_{\xi_i}(W \mid X = \widehat{X}_{*i})\}_{i=1}^m$. Note that, of course, the natural estimate of $Q_\tau(W)$ is not $\widehat{Q}_\tau(W)$, but the τth sample quantile of $\{W_i\}_{i=1}^n$; the former is of interest because it is coherent with the proposed model and it suggests how counterfactual quantiles can be estimated. Also note that this simulation-based estimation procedure is similar in spirit to (but not coincidental with) that of Melly (2005), though Melly (2006) shows if the size of the simulation draws m increased to infinity, then both would be numerically equivalent. Finally, note that Albrecht et al. (2009) show that $\widehat{Q}_\tau(W)$ is a consistent estimate of $Q_\tau(W)$, if consistent estimates of the conditional quantiles $Q_\upsilon(W \mid X = x)$ are used; in our context, this can be ensured under the regularity conditions specified in Lee (2007).

3.1 Counterfactual quantiles fixing all regressors or all returns

Hereafter, we use the superscript $t = r$ in operators, parameters, estimators and observations in order to emphasize that they are considered at time period r; thus the τth marginal quantile of W in period r will be denoted by $Q_\tau^{t=r}(W)$, and so on.

In equation (3.1) the τth marginal quantile of W in period r is defined implicitly as a solution to an equation which depends on the marginal distribution of X in period r and the conditional density of $W \mid X = x$ in period r. Then, given $r, r^* \in \{1, ..., T\}$, with $r^* \leq r$, if we were interested in the τth quantile of wages that would have prevailed in period r if the distribution of individual attributes had remained distributed as in period r^*, and workers had been paid according to the wage schedule observed in period r, what we should find is the solution q to a similar equation to (3.1), but using the marginal distribution of X in period r^* and the conditional density of $W \mid X = x$ in period r :

$$\int \left(\int 1\{w \leq q\} f_{W|X=x}^{t=r}(w)\mathrm{d}w \right) \mathrm{d}F_X^{t=r^*}(x) = \tau. \tag{3.2}$$

We denote this τth counterfactual quantile by $Q_\tau^{t_X=r^*,t_\theta=r}(W)$, where θ stands for the set of parameters that determine the returns to X through all quantiles, i.e. $\theta \equiv \{(\beta_\upsilon, \gamma_\upsilon', \lambda_\upsilon(\cdot))\}_{\upsilon \in (0,1)}$. Following the methodology of Machado and Mata (2005), an estimate of $Q_\tau^{t_X=r^*,t_\theta=r}(W)$, denoted by $\widehat{Q}_\tau^{t_X=r^*,t_\theta=r}(W)$, is the τth sample quantile of $\{\widehat{Q}_{\xi_i}^{t=r}(W \mid X = \widehat{X}_{*i}^{t=r^*})\}_{i=1}^m$, where $\{\xi_i\}_{i=1}^m$ is a draw of size m from a uniform distribution on $(0,1)$, $\{\widehat{X}_{*i}^{t=r^*}\}_{i=1}^m$ is a draw of size m with replacement from $\{\widehat{X}_i^{t=r^*}\}_{i=1}^{N_{r^*}}$, $\widehat{X}_i^{t=r^*} \equiv (S_i^{t=r*}, (Z_{1i}^{t=r^*})', \widehat{V}_i^{t=r*})$ and $\widehat{V}_i^{t=r*} \equiv S_i^{t=r^*} - \widehat{\mu}^{t=r^*} - (\widehat{\pi}^{t=r^*})'Z_i^{t=r^*}$. Note that $\widehat{Q}_{\xi_i}^{t=r}(W \mid X = \cdot)$ is the estimator defined in (2.9) computed with the observations in $t = r$, but then this function is evaluated at observations of the individual characteristics in $t = r^*$.

Once the τth counterfactual quantile is estimated, a decomposition of the difference between the estimated τth marginal quantiles in periods r and r^* may be proposed:

$$\widehat{Q}_\tau^{t=r}(W) - \widehat{Q}_\tau^{t=r^*}(W) = \left[\widehat{Q}_\tau^{t_X=r,t_\theta=r}(W) - \widehat{Q}_\tau^{t_X=r^*,t_\theta=r}(W)\right] + \tag{3.3}$$
$$\left[\widehat{Q}_\tau^{t_X=r^*,t_\theta=r}(W) - \widehat{Q}_\tau^{t_X=r^*,t_\theta=r^*}(W)\right].$$

The first term on the right-hand side is the part of the difference between quantiles that is explained by changes in the distribution of the individual characteristics between periods r^* and r, and is usually referred to as the composition effect. The second term is the part of the difference that is explained by changes in the returns to the individual characteristics.

A counterfactual quantile which assumes that returns are kept fixed at $t = r^*$ can be defined in a similar way as $Q_\tau^{t_X=r^*,t_\theta=r}(W)$, by simply exchanging the roles of r and r^* in the previous definitions. If we denote the estimate of this counterfactual quantile by $\widehat{Q}_\tau^{t_X=r,t_\theta=r^*}(W)$, then an alternative decomposition to (3.3) may be proposed using the differences $[\widehat{Q}_\tau^{t_X=r,t_\theta=r}(W) - \widehat{Q}_\tau^{t_X=r,t_\theta=r^*}(W)]$ and $[\widehat{Q}_\tau^{t_X=r,t_\theta=r^*}(W) - \widehat{Q}_\tau^{t_X=r^*,t_\theta=r^*}(W)]$. There is no reason why this alternative decomposition and (3.3) should lead to similar results, i.e. the order in which the decomposition is made may matter.

3.2 Counterfactual quantiles fixing some (but not all) regressors

The methodology that is used in the previous subsection also allows us to define and estimate counterfactual quantiles keeping fixed only the distribution of some of the explanatory variables of the model. In this case, several alternative counterfactual distributions can be defined, and one must choose in each application which one is the most appropriate or relevant for the analysis to be performed. Hereafter we consider that X is divided into two subvectors $X_1 \in \mathbb{R}^{d_1}$ and $X_2 \in \mathbb{R}^{d_2}$.

First, let us assume that we are interested in the distribution of wages that would have prevailed if the individual characteristics X_1 had remained distributed as in period r^*, but workers had been paid according to the wage schedule observed in period r and the remaining individual attributes X_2 are distributed as in period r. Then, the τth counterfactual quantile of this wage distribution, denoted by $Q_\tau^{t_{X_1}=r^*,t_{X_2|X_1}=r,t_\theta=r}(W)$, is defined implicitly as the solution q to the equation

$$\int \left[\int \left(\int 1\{w \le q\} f_{W|X=x}^{t=r}(w)\mathrm{d}w\right) \mathrm{d}F_{X_2|X_1=x_1}^{t=r}(x_2)\right] \mathrm{d}F_{X_1}^{t=r^*}(x_1) = \tau. \tag{3.4}$$

When estimating this counterfactual quantile, it is necessary to generate samples from an estimate of $F_{X_2|X_1=x_1}^{t=r}(x_2)$. We propose to use the Nadaraya-Watson estimate $\widehat{F}_{X_2|\hat{X}_1=x_1}^{t=r}(x_2) \equiv$

$\sum_{j=1}^{N_r} 1\{\widehat{X}_{2j}^{t=r} \leq x_2\}\omega_{\widehat{X}_{1,j}}^{t=r}(x_1)$, where $\omega_{\widehat{X}_{1,j}}^{t=r}(x_1) \equiv K(h_1^{-1}(x_{11} - \widehat{X}_{11j}^{t=r}),...,h_{d_1}^{-1}(x_{1d_1} - \widehat{X}_{1d_1j}^{t=r}))/$
$\sum_{l=1}^{N_r} K(h_1^{-1}(x_{11} - \widehat{X}_{11l}^{t=r}),...,h_{d_1}^{-1}(x_{1d_1} - \widehat{X}_{1d_1l}^{t=r}))$; $K(\cdot)$ is a d_1-dimensional kernel function; h_1,
..., h_{d_1} are bandwidths and $\widehat{X}_{1sj}$ denotes the sth component of $\widehat{X}_{1j}$. Note that a draw of
$\widehat{F}_{X_2|\widehat{X}_1=x_1}^{t=r}(x_2)$ is $\widehat{X}_{2j_*}^{t=r}$, where j_* is a draw from a discrete random variable J with proba-
bility mass function $P(J = j) = \omega_{\widehat{X}_{1,j}}^{t=r}(x_1)$ for $j \in \{1,...,N_r\}$. As a consequence, an esti-
mate of $Q_\tau^{t_{X_1}=r^*,t_{X_2|X_1}=r,t_\theta=r}(W)$, denoted by $\widehat{Q}_\tau^{t_{X_1}=r^*,t_{X_2|X_1}=r,t_\theta=r}(W)$, can be found as fol-
lows: i) obtain a draw of size m from the uniform distribution on $(0,1)$, denoted as $\{\xi_i\}_{i=1}^m$;
ii) obtain a draw of size m with replacement from $\{\widehat{X}_{1i}^{t=r^*}\}_{i=1}^{N_{r^*}}$, denoted as $\{\widehat{X}_{*1i}^{t=r^*}\}_{i=1}^m$; iii) for
$i \in \{1,...,m\}$, obtain one draw from a discrete random variable J with probability mass function
$P(J = j) = \omega_{\widehat{X}_{1,j}}^{t=r}(\widehat{X}_{*1i}^{t=r^*})$ for $j \in \{1,...N_r\}$; this draw is denoted $j_*(i)$; iv) find the τth sample
quantile of $\{\widehat{Q}_{\xi_i}^{t=r}(W \mid X_1 = \widehat{X}_{*1i}^{t=r^*}, X_2 = \widehat{X}_{2j_*(i)}^{t=r})\}_{i=1}^m$. Note that this estimator is based on a
nonparametric estimator of a conditional distribution given X_1 and, hence, it is affected by the
curse of dimensionality if d_1 is large; this problem could be circumvented by assuming a paramet-
ric specification for the conditional distribution of $X_2 \mid X_1 = x_1$. Also note that in all definitions
of this subsections, returns could also be fixed at the same period as the marginal CDF of X_1,
by simply using $f_{W|X=x}^{t=r^*}(\cdot)$ instead of $f_{W|X=x}^{t=r}(\cdot)$ in (3.4), and $\widehat{Q}_{\xi_i}^{t=r^*}(W \mid X_1 = \cdot, X_2 = \cdot)$ instead
of $\widehat{Q}_{\xi_i}^{t=r}(W \mid X_1 = \cdot, X_2 = \cdot)$ in the fourth step above.

The estimator of the counterfactual quantiles that we propose differs from the one introduced
in Machado and Mata (2005) when the distribution of some of the regressors are held fixed.
Machado and Mata (2005) derive an estimator where the support of X_1 is divided into a finite
number of partitions; hence the difference between our estimator and theirs is the same as the
difference between an histogram and a nonparametric kernel estimate. We have preferred to use
nonparametric kernel estimates because it has better statistical properties, and the problems
of bandwidth selection and kernel choice have been studied more in detail than the problem
of selecting partitions in histogram-like estimation. Additionally, the estimator we propose is
easier to implement in the case of multivariate X_1.

Once this τth counterfactual quantile is estimated, the composition effect the first term in
(3.3) may be further decomposed as follows:

$$Q_\tau^{t_X=r,t_\theta=r}(W) - \widehat{Q}_\tau^{t_X=r^*,t_\theta=r}(W) = \tag{3.5}$$
$$= \left[\widehat{Q}_\tau^{t_X=r,t_\theta=r}(W) - \widehat{Q}_\tau^{t_{X_1}=r^*,t_{X_2|X_1}=r,t_\theta=r}(W)\right] +$$
$$\left[\widehat{Q}_\tau^{t_{X_1}=r^*,t_{X_2|X_1}=r,t_\theta=r}(W) - \widehat{Q}_\tau^{t_X=r^*,t_\theta=r}(W)\right].$$

The first term on the right-hand side is the part of the composition effect that is explained by
changes in the marginal distribution of the individual characteristics X_1, and the second term
is the part of the composition effect that is explained by changes in the conditional distribution

of $X_2 \mid X_1$.

If the dimension of X is greater than 2, then there are various possible ways to split the explanatory variables into X_1 and X_2, and hence various different decompositions in the spirit of (3.5) are possible. In addition, each term on the right-hand side of (3.5) can be further decomposed, using an additional decomposition of X_1 or X_2, and the same methodology as above. Among all these possible decompositions of the composition effect, the researcher should decide which ones are of interest to her.

3.3 Counterfactual quantiles fixing some (but not all) returns

The previous methodology also allows us to define counterfactual quantiles keeping fixed only some of the returns of the model. Here we just present some of the counterfactual quantiles that can be defined; with the same methodology many others can be defined.

First, let us assume that we are interested in the τth counterfactual quantile of wages that would have prevailed with individual attributes distributed as in period r^*, returns to schooling as in period r^* and all other returns as in period r. This counterfactual quantile, denoted by $Q_\tau^{t_X=r^*, t_\beta=r^*, t_{(\gamma,\lambda)}=r}(W)$, can be defined as the solution q to an equation similar to (3.2), but replacing $f_{W|X=x}^{t=r}(w)$ with $f_{W|X=x}^{t_\beta=r^*, t_{(\gamma,\lambda)}=r}(w)$, which denotes the PDF derived from the quantile function $Q_\tau^{t_\beta=r^*, t_{(\gamma,\lambda)}=r}(W \mid X = x) \equiv \beta_\tau^{t=r^*} s + z_1' \gamma_\tau^{t=r} + \lambda_\tau^{t=r}(v)$. And this counterfactual quantile can be estimated by $\widehat{Q}_\tau^{t_X=r^*, t_\beta=r^*, t_{(\gamma,\lambda)}=r}(W)$, which is defined in the same way as $\widehat{Q}_\tau^{t_X=r^*, t_\theta=r}(W)$, but replacing $\widehat{Q}_{\xi_i}^{t=r}(W \mid X = x)$ by $\widehat{Q}_{\xi_i}^{t_\beta=r^*, t_{(\gamma,\lambda)}=r}(W \mid X = x) \equiv \widehat{\beta}_{\xi_i}^{t=r^*} s + z_1' \widehat{\gamma}_{\xi_i}^{t=r} + \widehat{\lambda}_{\xi_i}^{t=r}(v)$. In a similar fashion, it is possible to define other counterfactual quantiles considering some returns in period r^* and the remaining ones in period r. Once these counterfactual quantiles are estimated, the difference between the marginal quantiles in periods r and r^* that is explained by changes in returns, i.e. the second term on the right-hand side of (3.3), can be further decomposed, for instance, as follows:

$$
\begin{aligned}
\widehat{Q}_\tau^{t_X=r^*, t_\theta=r}(W) &- \widehat{Q}_\tau^{t_X=r^*, t_\theta=r^*}(W) = \\
&= \left[\widehat{Q}_\tau^{t_X=r^*, t_\theta=r}(W) - \widehat{Q}_\tau^{t_X=r^*, t_\beta=r^*, t_{(\gamma,\lambda)}=r}(W)\right] + \\
&\quad \left[\widehat{Q}_\tau^{t_X=r^*, t_\beta=r^*, t_{(\gamma,\lambda)}=r}(W) - \widehat{Q}_\tau^{t_X=r^*, t_{(\beta,\lambda)}=r^*, t_\gamma=r}(W)\right] + \\
&\quad \left[\widehat{Q}_\tau^{t_X=r^*, t_{(\beta,\lambda)}=r^*, t_\gamma=r}(W) - \widehat{Q}_\tau^{t_X=r^*, t_\theta=r^*}(W)\right].
\end{aligned}
\tag{3.6}
$$

Here, the first term on the right-hand side is the part of the difference that is explained by changes in the returns to schooling, the second term is the part of the difference that is explained by changes in the labour market returns to the unobservables affecting schooling, and the third term is the part of the difference that is explained by changes in the returns to the observable

control variables in Z_1.

Another type of counterfactual exercise may be performed in order to disentangle the effect of changes in one of the returns, say the median ones $\theta_{0.5} \equiv (\beta_{0.5}, \gamma'_{0.5}, \lambda_{0.5}(\cdot))'$, from the effect of changes in the corresponding errors, say the median ones $\varepsilon_{0.5} \equiv U_{0.5} - \lambda_{0.5}(V_\alpha)$, using the same methodology as in Melly (2005, p.581). This allows one to decompose the difference between the marginal quantiles in periods r and r^* that is explained by changes in returns into that part that is explained by the changes in median returns and that part that is explained by changes in the conditional distribution of the unobservable component of the median wage regression given X, or equivalently, by changes in the within-group or residual inequality.

4 Data

The requirement of finding instruments for schooling restricts the data sets that can be used to apply our methodology. The Panel Study of Income Dynamics (PSID) is an ideal data set for this study since it contains both information on the county of residence during the adolescence and also information on parental education.

We use the 1983 and 1993 waves of the PSID. In both waves we restrict our attention to current male heads in each period who report to work for someone else. These individuals were members of the original families chosen in 1968 and, hence, our dataset does not take into account the immigration effects from 1968 onwards; to overcome this problem, several samples with immigrants were added to the PSID in the mid-nineties, but they were not available for the period considered in our analysis. In both waves we only consider individuals from the Survey Research Center (SRC) and discard those individuals belonging to the Census Sample (which was added to the PSID cross-section in order to oversample low-income families and it would decrease representativeness of the original PSID sample if included). We choose the waves for 1983 and 1993 in order to capture the increase in wage inequality experienced during the eighties in the United States (see e.g. Levy and Murnane 1992); additionally, our instrument is only available for this time window. We do not exploit the panel dimension and consider these two waves as two cross sections. The exact definition of wages, schooling and the variables included in Z can be found in the Appendix.

For our study we only consider male heads between 22 and 65 years old for whom the two variables that are used to compute hourly wages are positive. Younger workers are not considered to ensure that all workers in the sample have possibly completed their schooling decisions. Similar results are obtained if male heads between 18 and 21 years old are included. In addition, all observations with some missing value for any of the regressors are also excluded.

The final sample sizes are 1469 individuals for 1983 and 1633 individuals for 1993. In Table 1 we report descriptive statistics of all the variables. Note that hourly wages increased, on average, by 2 dollars per hour; however, the median of the distribution remained unchanged (10.94 in 1983 and 10.97 in 1993), the 95th quantile increased (22.90 in 1983 and 29.18 in 1993) and the 5th quantile decreased (4.09 in 1983 and 3.92 in 1993). There is also evidence that the ratio between quantiles at the top of the distribution increased (the ratio between the 90th quantile and the 70th quantile is 1.40 in 1983 and 1.52 in 1993), whereas the ratio between the quantiles at the bottom of the distribution slightly decreased (the ratio between the 30th quantile and the 10th quantile is 1.59 in 1983 and 1.57 in 1993). All these results show that wage inequality increased between 1983 and 1993. To better understand the sources of this increase in wage inequality, it is important to note the increase in education attainment: the number of college graduates and postgraduates increased by 5% and the percentage of individuals with only high school or dropouts slightly decreased; also the distribution of parental education shifted to the right between both periods.

The vector Z to be used as a set of regressors in the schooling equation must include the exogenous variables in the wage equation Z_1 and valid instruments Z_2 excluded from the wage equation. In order to be able to quantify the causal relationship between schooling and earnings, an exogenous source of variation in education outcomes is required. Much attention has been focussed on the institutional sources of variation in schooling, arising from minimum school leaving age, tuition costs for higher education or from the geographic proximity to schools. Following Card (1995), Carneiro and Lee (2009) and Kling (2001), among others, we consider proximity to college as a measure of these institutional factors. This variable aims to assess the cost of attending schooling based on the geographical variation of the individuals when they were young: those individuals who grow up in a county without a college face a higher cost of continuing education and, therefore, there is a decrease in the level of schooling that is eventually achieved. The choice of this instrument is not free of controversy; Carneiro and Heckman (2002) showed that distance to college is correlated with the result of a test score capturing cognitive ability in the National Longitudinal Survey of Youth; since we do not have access to this proxy for ability in the PSID, distance to college is the most appropriate instrument at hand. We introduce proximity to college in our model considering a dummy variable denoted by *proximity* that takes value 1 if there is at least a two-year post-secondary institution within the county where the individual grew up and 0 otherwise; this variable can be constructed by combining data on the PSID with the information on the location and foundation year of all public two-year and four-year colleges in the United States reported by the National Center of

Table 1: Descriptives of main variables. Mean of each variable with standard errors in brackets. For education, percentage of observations in each level with standard errors in brackets

	Year 1983 (1469 obs.)		Year 1993 (1633 obs.)	
Total wages	24868.79 (15141.63)		27507 (22648.32)	
Hourly wages	12.1885 (17.7681)		13.861 (14.504)	
Age	36.8944 (10.8846)		38.55 (9.82)	
Tenure	12.3474 (9.7633)		13.7440 (8.6924)	
Size of City Dummy	0.5207 (0.4997)		0.4868 (0.4999)	
Govern. Occup.	0.1912 (0.3934)		0.1984 (0.3989)	
White Dummy	0.9278 (0.2588)		0.9326 (0.2507)	
South	0.3036 (0.4599)		0.3129 (0.4638)	
North	0.3158 (0.4650)		0.3312 (0.4708)	
West	0.1803 (0.3846)		0.1592 (0.3659)	
Northeast	0.2001 (0.4002)		0.1965 (0.3975)	
Proximity to College	0.7243 (0.4470)		0.7495 (0.4334)	
Industry Dummies				
Agriculture & Fisheries	0.0231 (0.1504)		0.0202 (0.1407)	
Mining	0.0149 (0.1214)		0.0091 (0.0954)	
Construction	0.0769 (0.1665)		0.0857 (0.2800)	
Manufacturing	0.3042 (0.4602)		0.2584 (0.4379)	
Transport &Communication	0.1266 (0.3326)		0.1065 (0.3086)	
Wholesale/Retail Trade	0.1463 (0.3535)		0.1671 (0.3732)	
Finance, Insurance, Real State	0.0394 (0.1948)		0.0404 (0.1969)	
Business & Repair Serv.	0.0408 (0.1979)		0.0545 (0.2270)	
Personal Serv.	0.0088 (0.0936)		0.0085 (0.0922)	
Entertainment/Recreation Serv.	0.0061 (0.0780)		0.0091 (0.0954)	
Professional Serv.	0.1198 (0.3248)		0.1439 (0.3511)	
Public Admin.	0.0925 (0.2899)		0.0961 (0.2948)	
Education		% Close to College		% Close to College
% < High School ($S < 12$)	13.75 (0.3445)	57.92	10.90 (0.3117)	67.97
% High School ($S = 12$)	38.39 (0.4865)	71.45	34.41 (0.4752)	73.67
% 1 Year College ($S = 13$)	6.73 (0.2507)	80.81	8.26 (0.2754)	74.81
% 2 Year College ($S = 14$)	11.70 (0.3216)	76.16	10.90 (0.3117)	80.34
% 3 Year College ($S = 15$)	3.13 (0.1742)	78.26	4.16 (0.1998)	76.47
% College Graduate ($S = 16$)	15.72 (0.3641)	76.62	19.65 (0.3975)	79.13
% Postgraduate ($S = 17$)	10.55 (0.3073)	77.42	11.69 (0.3214)	72.77
Father's Education				
% < High School (F1,F2,F3)	51.05 (0.5000)	64.93	36.25 (0.4808)	67.06
% High School (F4,F5)	29.27 (0.4551)	77.21	37.78 (0.4849)	75.36
% Some College (F6)	7.69 (0.2665)	78.76	8.81 (0.2836)	80.56
% College (F7,F8)	11.43 (0.3183)	89.29	16.65 (0.3727)	88.60
Mother's Education				
% < High School (M1,M2,M3)	38.36 (0.4876)	63.05	25.78 (0.4375)	66.27
% High School (M4,M5)	45.88 (0.4984)	68.43	53.52 (0.4989)	75.97
% Some College (M6)	7.01 (0.2554)	82.52	9.92 (0.2990)	82.72
% College (M7,M8)	8.03 (0.2718)	82.20	10.71 (0.3094)	83.43

Education Statistics (NCES) in the Higher Education General Information Survey (HEGIS), for the years 1966 to 1985, and in the Integrated Postsecondary Education Data System (IPEDS) (Institutional Characteristics Survey) from 1986 onwards; for each institution we use the year when it was founded (in fact, we use the first year a 2-year or a 4-year degree was offered), and by matching this information with the county of residence where the individual grew up, we determine the presence of a two-year or a four-year postsecondary institution. We assume that this information provided by the PSID coincides with the county where the individual lived when he was 16. As opposed to Card (1995), the 2-year degrees have been considered important in the college decision in Carneiro and Lee (2009) and Kling (2001). However, as a robustness check, first stage quantile estimates for the schooling equation were obtained using in the definition of proximity to college only those institutions offering 4-year degrees in 1993. We obtain that the difference of these results with the results reported in the remainder of the paper are significantly small (e.g. the average difference between the estimated schooling coefficient over 40 equally spaced quantiles is 0.0089 in 1983 and -0.0011 in 1993, with bootstrapped standard errors for these differences of 0.0191 in 1983 and 0.0227 in 1993).

In Table 1 we report, for each education group, the percentage of individuals for which there is at least a 2-year college in the county where he grew up. In both years, the percentage of college graduates that lived near a college is significantly greater than the percentage of dropouts living close to a college. It has also been discussed in the literature that the effect on schooling decisions of a nearby college differs with family background. Thus, the impact on schooling of decreasing the cost of higher education for those individuals close to a college must be greater for poor families or families with low educated parents. In order to control for this, Z_2 includes, in addition to *proximity*, interaction variables between proximity to college and 14 dummies for parental education. The identifying assumption is that, controlling for individual and parental education, college proximity should have no direct effect on wage for any parental background group. It is important to control for parental education in the earnings and schooling equations in order to ensure that this identifying assumption holds, since otherwise distance to college could be correlated with wages and schooling decisions through parental education. This correlation is observed in Table 1: the percentage of dropout parents of those individuals living near a college is almost 20 points lower than the percentage of graduated parents of those individuals living close to a college.

Finally, it is important to note that other instruments for schooling that have been used in previous literature were not available for our sample or proved to be non-significant. More specifically, following Angrist and Krueger (1991), we could have derived an instrument for

schooling by using the exogenous variation implied by the quarter of birth in the amount of years of education attained; but, given our small sample size, only a few observations per year and per quarter are available and, in our samples, those individuals who were born earlier in the year do not achieve lower education level than those who were born later in the year. On the other hand, following Carneiro and Lee (2009) among others, average county or state tuition fees could also be used to approximate the cost of attending college; the data from HEGIS and IPEDS only contain information on tuition fees from 1969 onwards and, therefore, for most of the individuals in our samples it is not possible to derive the average tuition fee in the county or state where he grew up. With the information that is available, we derived a proxy for average county tuition fees, assuming that the tuition fees in years previous to 1969 were the same as in 1969; the results that are obtained including it in Z_2 are similar to the ones that are reported below.

5   Results

5.1   Estimation of the Wage and Schooling Equations

We estimate the model defined by (2.1)-(2.3) with different values of τ for both periods 1983 and 1993. For comparative purposes, we also provide the estimates that are obtained when schooling S is assumed to be exogenous, i.e. assuming that the first equation in (2.1) holds with an intercept term and error term satisfying that $Q_\tau \left(U_\tau \mid S = s, Z_1 = z_1 \right) = 0$ for any $(s, z_1')'$ in the support of $(S, Z_1')'$; these results will be referred to as results for the "exogenous model", whereas those obtained for the model defined by (2.1)-(2.3) will be referred to as results for the "endogenous model".

The upper panel in Table 2 reports the estimated coefficients of the instrumental variables Z_2 in the first stage estimation. The fact of having grown up in a county with a college has a significant positive effect on the schooling level of those individuals whose parents only achieved the lowest educational level. For this group, the median schooling for those individuals living close to a college is 2.5 years higher in 1983 and 1.7 years higher in 1993 than for those individuals that lived further away. The effect of distance to college is monotonically decreasing with the education of the father; in fact, we obtain that in 1983 distance to college has almost no impact for those individuals whose fathers finished college. Maternal education does not have any differential effect on the impact of distance to college in 1993; however, in 1983 living close to a college increases schooling by 1 year for those individuals with mother's education above the sixth grade. The bottom panel in Table 2 reports the p-values of the joint significance test of the

Table 2: Estimates of the coefficients of the instrumental variables Z_2 in the first stage estimation of the conditional median (upper panel) and Test of joint significance of regressors in Z_2 (lower panel).

	Year 1983			Year 1993		
Dep. Variable: Years of Schooling	Coeff.		p-value	Coeff.		p-value
proximity to college	2.548198**		0.000	1.737971*		0.014
proximity*father education 2	-.9891704*		0.038	-1.115219*		0.055
proximity*father education 3	-.9551701**		0.099	-1.517416*		0.022
proximity*father education 4	-1.201334*		0.023	-1.244606*		0.040
proximity*father education 5	.6578266		0.473	-1.014563		0.296
proximity*father education 6	-1.362265*		0.032	-2.315919**		0.001
proximity*father education 7	-.2193215		0.762	-.7503964		0.309
proximity*father education 8	-2.457444*		0.014	-1.382607		0.122
proximity*mother education 2	-1.60994**		0.001	-.3863		0.590
proximity*mother education 3	-1.645981**		0.004	-0.7953959		0.292
proximity*mother education 4	-1.757884**		0.001	-.6746985		0.345
proximity*mother education 5	-1.084456		0.247	1.505461		0.112
proximity*mother education 6	-1.593101*		0.017	-.286328		0.722
proximity*mother education 7	-1.792687*		0.016	-1.08793		0.199
proximity*mother education 8	-1.084995		0.278	0.3964252		0.693
	H_0 : coeff of $Z_2 = 0$					
first-stage quantile =0.5	$F(15,1421)=2.62$	pvalue	0.0007	$F(15,1585)=2.28$	pvalue	0.0035
first-stage quantile =0.05	$F(15,1421)=5.12$	pvalue	0.0000	$F(15,1585)=2.80$	pvalue	0.0003
first-stage quantile =0.25	$F(15,1421)=3.82$	pvalue	0.0000	$F(15,1585)=3.43$	pvalue	0.0000
first-stage quantile =0.75	$F(15,1421)=1.15$	pvalue	0.3034	$F(15,1585)=3.05$	pvalue	0.0001
first-stage quantile =0.95	$F(15,1421)=44.45$	pvalue	0.0000	$F(15,1585)=6.65$	pvalue	0.0000

Note: First stage median regression additionally includes as regressors an intercept and vector Z_1 (tenure, squared tenure, race dummy, dummy for size of city, dummy for government occupation, regional dummies, dummies for parental education and industry dummies. Significant estimates at 1% significance level are denoted **; significant estimates at 5% significance level are denoted *.

excluded variables Z_2 when estimating the quantile model $Q_\alpha(S \mid Z) = \mu_\alpha + \pi'_\alpha Z$, for various α; a joint significant impact of distance to college and its interactions with parental education is found for most of the schooling quantiles, except for some quantiles at the top. Using these estimates for various quantiles α, the test of Koenker and Basset (1982) rejects the independence of V and Z. Therefore, it is not immediate to determine which first stage quantile is optimal one in terms of efficiency as discussed in the Estimation Section and the median as the first stage quantile is assumed. Following Card (1995), the interpretation that the reduction in the cost of education of a close college is higher for families with poor background allows us to check whether distance to college is a valid instrument. There are reasons to suspect that individuals who grew up near a college may have higher earnings after controlling for parental education and current geographical information. Under the identifying assumption that the direct earning effects of having lived near a college does not differ by family or parental background, we are able to use the interactions of distance to college and parental education as instruments, and use the distance to college dummy as a direct control in the schooling equation. The results of this quantile regression allow us to conclude that distance to college is not a significant variable in the wage equation and therefore these estimates provide no evidence against the assumption that distance to college provides an exogenous variation of schooling.

Before estimating the conditional quantiles of wages, we have estimated its conditional mean. The OLS and IV estimates (and standard errors) of the education coefficient are, respectively, 0.0764 (0.00595) and 0.0464 (0.05722) for year 1983, and 0.0814 (0.00766) and 0.0634 (0.07231) for year 1993. In both periods, the IV estimates are below the corresponding OLS estimate, as expected in a model with homogenous returns to schooling and positive correlation between schooling and innate ability (see e.g. Carneiro and Heckman, 2002).

Figure 1 shows the quantile regression estimates for the education coefficient β_τ for various τ in both periods; the third graph in this figure plots the changes in the coefficients between the two years. If education is endogenous, then the estimates of β_τ obtained with the exogenous model are misleading and biased. The first feature to notice when comparing the estimates of β_τ from the exogenous and the endogenous model is that the former are roughly constant around 0.06 and 0.09, whereas the latter show noteworthy differences across quantiles. The third graph in Figure 1 also shows that the changes in the quantile regression estimates between both years have been more stable in the exogenous model than in the endogenous model, although for the latter the difference is not statistically significant. The differences in the adjusted-by-endogeneity returns to education arise mainly because of the increase in the returns for the individuals at the bottom of the distribution, and the decrease of the returns for those individuals at the top of
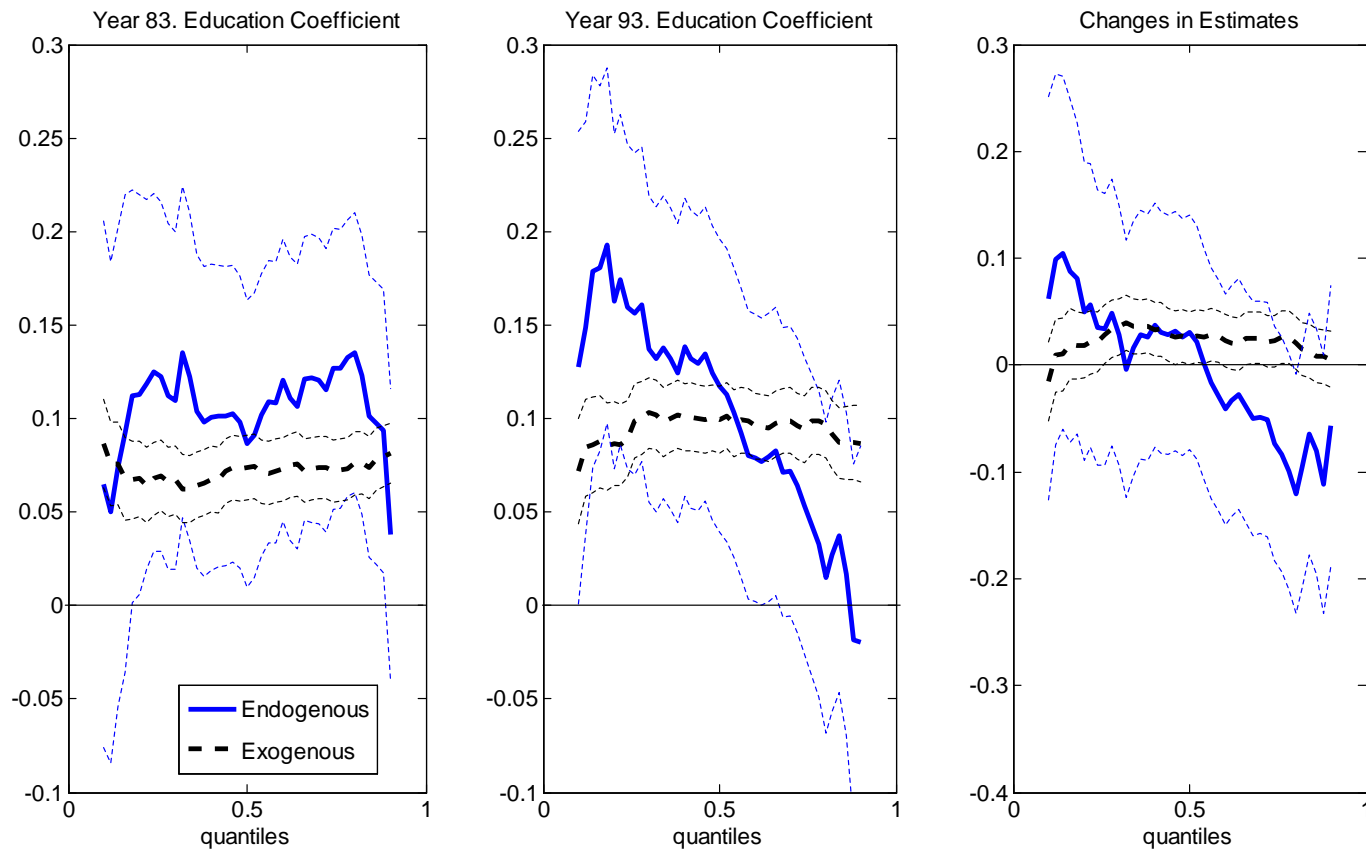
Figure 1: Estimates of the coefficients of education with the exogenous and the endogenous model for various quantiles. Dotted lines represent 95% pointwise confidence intervals derived from the asymptotic distribution by Lee (2007). Note: A fifth order polynomial was used to approximate function $\lambda_\tau(\cdot)$ in the second-stage estimation, i.e. $p_k(s) = s^{k-1}$, $k_N = 6$, and in the trimming function we use $[\tilde{Q}_{0.02}(S), \tilde{Q}_{0.98}(S)]$ as A_S, $[\tilde{Q}_{0.02}(\hat{V}), \tilde{Q}_{0.98}(\hat{V})]$ as A_V, and the product of the intervals $[\tilde{Q}_{0.02}(Z_{1j}), \tilde{Q}_{0.98}(Z_{1j})]$, for $j = 1, ..., p_1$, as A_{Z_1}, (where $\tilde{Q}_\tau(\xi)$ denotes the τth sample quantile of $\{\xi_i\}_{i=1}^N$). The bandwidth choice is the product of the standard deviation of the residuals of each linear quantile regression and $N^{-0.15}$; this bandwidth satisfies the asymptotic restrictions imposed in Lee (2007) if $\lambda_\tau(\cdot)$ is at least five times continuously differentiable.

the wage distribution. The finding that quantile effects are larger at the lower end of the wage distribution in the adjusted-by-endogeneity returns is also found in Lee (2007) and Honoré and Hu (2004) for the 1980 US Census using different instruments; however, the opposite is found in Buchinsky (1994), Machado and Mata (2005) and Melly (2005), who do not control for the endogeneity of education. The differences between the estimated coefficients for the variables in Z_1 with both models are less important than those observed for the coefficient of education. The difference between the estimated coefficients with the exogenous and the endogenous models can be explained using the omitted variables bias formula derived in Angrist et al. (2006).

In Figure 2, we plot the estimates of the nonparametric function $\lambda_\tau(\cdot)$ that are obtained in the third stage estimation of the conditional quantiles, for various τ. This figure can be interpreted as the quantile regression equivalent of the Hausman test for endogeneity in the mean regression, where the statistical significance of the first-stage residuals is tested in the wage equation; this figure shows evidence in favour of the existence of endogeneity, since it shows that the nonparametric control functions $\lambda_\tau(\cdot)$ are not constant over the support of the unobservables V.

5.2 Decomposition of Changes in the Wage Distribution

First of all, Table 3 reports the sample marginal quantiles of hourly wages (in logs), and the estimates of the marginal quantiles that are consistent with our model (described at the beginning of Section 3), and those that would be consistent with a model with exogenous schooling. In all cases, the three estimates are very similar. If schooling is endogenous and our model is correctly specified, then the results of the endogenous model provide a consistent estimate of the conditional quantile function $Q_\tau(W \mid S, Z_1, V)$ for $\tau \in (0,1)$. Since our data supports the assumption of endogeneity of schooling, the conditional quantile function $Q_\tau(W \mid S, Z_1)$ is not necessarily linear in S and Z_1. Therefore, although the estimates of the exogenous model provide the best linear prediction of this conditional quantile function (Angrist et al. 2006), it does not necessarily provide a consistent estimate of it. However, the results in Table 3 suggest that the approximation error between the conditional quantile function and its linear prediction must be very small. The similarities between the estimates of the endogenous and the exogenous model hold despite the fact that the coefficient for schooling in the latter is contaminated by the omission of V in the wage equation.
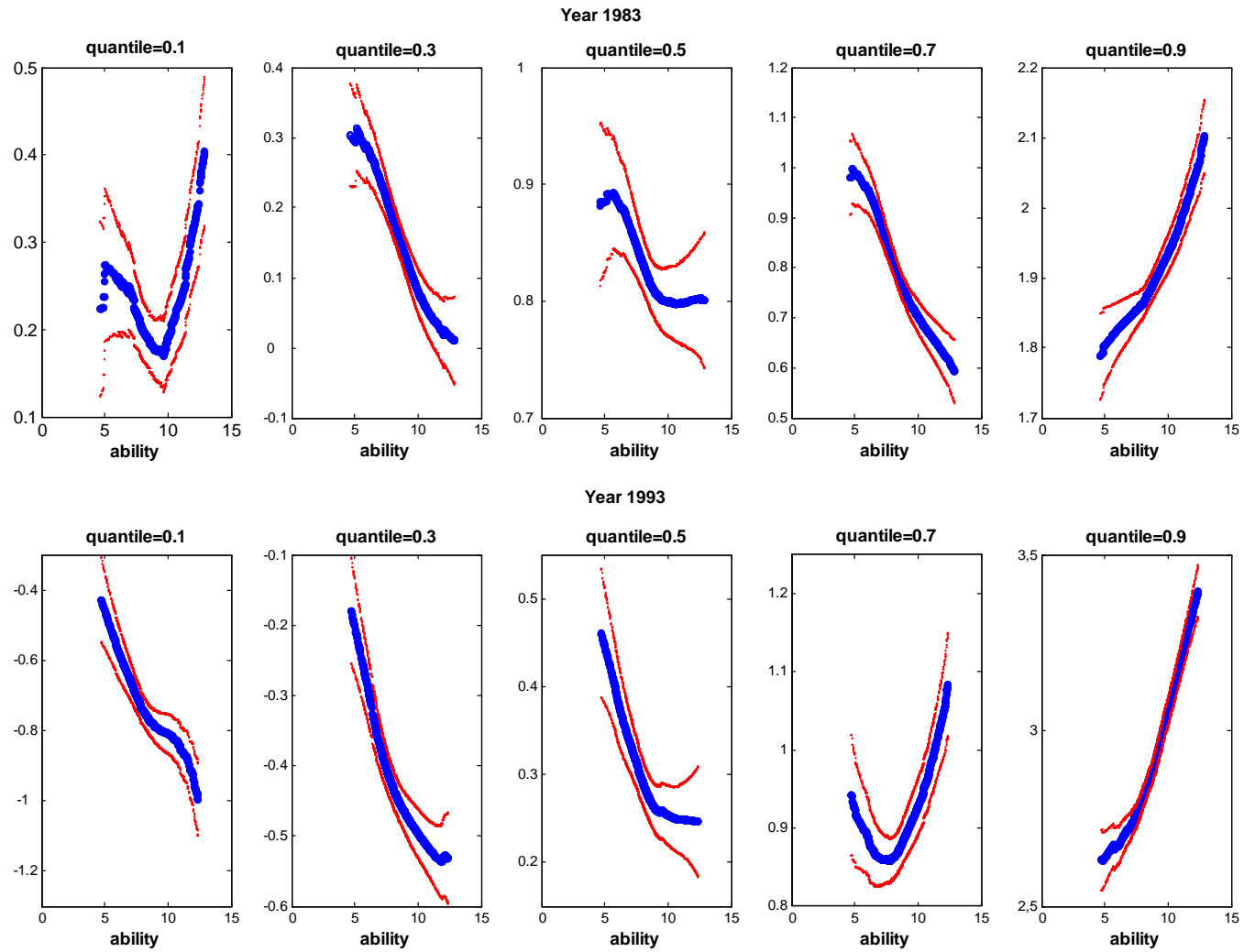
Figure 2: Local linear quantile regression estimates for $\lambda_\tau(\bullet)$ over the support of V for different quantiles τ with 95% confidence bands.
Note: A Gaussian kernel and a bandwidth selected by the rule of thumb given in Fan and Gijbels (1996) are used; confidence intervals were computed using the formula given in Chaudhuri (1991) for the homoskedastic case.

5.2.1 Decomposition of differences keeping all regressors fixed

Table 4 shows the decomposition of the changes in the quantiles of the wage distribution into changes that can be attributed to changes in the joint distribution of the individual attributes and changes in the returns of these characteristics as in (3.3). It should be pointed out that all counterfactual distributions are computed using the second-stage estimator of $\lambda_\tau (\cdot)$, for computational reasons. We performed robustness checks in order to ensure that the second-stage and the third-stage estimators of $\lambda_\tau (\cdot)$ yield very similar quantitative results. The second line in each cell reports the 95% confidence interval of the differences of estimated quantiles (derived with 100 bootstrap replications). We conjecture that a bootstrap procedure can be applied in order to compute the standard error of the counterfactual quantiles and their differences. This is not guaranteed, however, in conditions established in Chernozhukov et al. (2009); they require $\sqrt{n}-$convergence of $\widehat{Q}_\tau(W \mid X = \widehat{X}_i)$ to $Q_\tau(W \mid X = X_i)$, which in the case of the nonparametric three-stage estimator proposed in (2.9) is not satisfied. Monte Carlo experiments could be used to provide evidence in favour of the conjecture we make here in order to compute bootstrapped standard errors, but they are beyond the scope of this paper. The estimates of the marginal quantiles of wages allow us to estimate inequality measures. Specifically, given $\tau, \tau' \in (0, 1)$ with $\tau > \tau'$, the interquartile range (IQR) of wages for period $t = r$ between quantiles τ and τ', denoted as $IQR^{t=r}_{\tau,\tau'}(W)$, can be estimated by $\widehat{IQR}^{t=r}_{\tau,\tau'}(W) = \widehat{Q}^{t=r}_\tau(W) - \widehat{Q}^{t=r}_{\tau'}(W)$. Counterfactual inequality measures are defined and estimated similarly.

The first column of Table 4 allows us to conclude that the marginal distribution of wages in 1993 stochastically dominates the marginal distribution for 1983, since the differences between quantiles are all positive. Wage inequality increased between both periods, as shown by the changes in the various IQR and the fact that the quantile differences are larger for the top quantiles: the top wage quantile and the IQR$_{0.9,0.1}$ increased by 12% between both time periods. The increase in wage inequality was significantly noteworthy at the top of the distribution. The second column of Table 4 shows that the counterfactual distribution of wages that would have prevailed if the individual attributes had been distributed as in 1983, but had been paid according to the wage structure in 1993, is dominated by the actual distribution of wages in 1993. This means that the change in the distribution of the regressors from 1983 to 1993 would have shifted the wage distribution to the right if the returns had remained as in 1993. This would have been mostly beneficial for individuals at the bottom of the distribution. The results in the third column imply that the changes in the returns go in the opposite direction: if the distribution of the regressors had been kept fixed at 1983, the changes in the returns to attributes experienced from 1983 to 1993 would have decreased wages at most of the quantiles, except at the top one;

Table 3: Sample quantiles of wages and quantiles estimated from the model

τ	Log Hourly Wages 1983			Log Hourly Wages 1993		
	Sample quantile	$Q_\tau^{t=1983}(W)$ Endogenous	$Q_\tau^{t=1983}(W)$ Exogenous	Sample Quantile	$Q_\tau^{t=1993}(W)$ Endogenous	$Q_\tau^{t=1993}(W)$ Exogenous
10th quantile	1.6297 [1.5627;1.6903]	1.6572 [1.6070;1.7073]	1.6551 [1.6041;1.7060]	1.6564 [1.5871;1.7084]	1.6581 [1.6049;1.7114]	1.6582 [1.6051;1.7112]
30th quantile	2.0952 [2.0605;2.1352]	2.0969 [2.0633;2.1306]	2.1002 [2.0663;2.1341]	2.1077 [2.0735;2.1541]	2.1235 [2.0881;2.1589]	2.1227 [2.0878;2.1577]
50th quantile	2.3920 [2.3546;2.4338]	2.3740 [2.3443;2.4037]	2.3730 [2.3437;2.4023]	2.3953 [2.3694;2.4278]	2.4040 [2.3712;2.4368]	2.4051 [2.3724;2.4378]
70th quantile	2.6194 [2.5916;2.6423]	2.6201 [2.5934;2.6469]	2.6191 [2.5923;2.6460]	2.6744 [2.6436;2.7105]	2.6755 [2.6430;2.7079]	2.6793 [2.6466;2.7119]
90th quantile	2.9543 [2.9031;2.9883]	2.9496 [2.9141;2,9850]	2.9511 [2.9170;2.9852]	3.0959 [3.0386;3.1298]	3.0786 [3.0294;3.1277]	3.0765 [3.0312;3.1218]

Note: 95% bootstrap confidence intervals (with 100 replications) are provided in squared brackets.

Table 4: Decomposition of the total changes in the marginal quantiles of wages for the endogenous model

Quantile τ	Total Change $\widehat{Q}_\tau^{t=93}(W) - \widehat{Q}_\tau^{t=83}(W)$	Changes in Regressors $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$	Changes in Returns $\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W) - \widehat{Q}_\tau^{t_X=83,t_\theta=83}(W)$
10th quantile	0.0009 [-0.0759;0.0777]	0.0774* [0.0263;0.1285]	-0.0765* [-0.1145;-0.0385]
30th quantile	0.0265 [-0.0244;0.0774]	0.0809* [0.0447;0.1171]	-0.0544* [-0.0774;-0.0314]
Median	0.0300 [-0.0166;0.0766]	0.0751* [0.0432;0.1071]	-0.0451* [-0.0687;-0.0215]
70th Quantile	0.0553* [0.0102;0.1004]	0.0745* [0.0432;0.1059]	-0.0192 [-0.0429;0.0045]
90th Quantile	0.1290* [0.0684;0.1896]	0.0568* [0.0156;0.0981]	0.0722* [0.0373;0.1071]
Quantiles τ,τ'	$\widehat{IQR}_{\tau,\tau'}^{t=93}(W) - \widehat{IQR}_{\tau,\tau'}^{t=83}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W) - \widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W) - \widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=83}(W)$
$\tau=0.9, \tau'=0.1$	0.1281* [0.0423;0.2139]	-0.0206 [-0.0768;0.0356]	0.1487* [0.0527;0.2446]
$\tau=0.5, \tau'=0.1$	0.0291 [-0.0325;0.0906]	-0.0023 [-0.0418;0.0372]	0.0314 [-0.0363;0.0990]
$\tau=0.9, \tau'=0.5$	0.0990* [0.0415;0.1565]	-0.0183 [-0.0522;0.0156]	0.1173* [0.0571;0.1775]

Note: 95% bootstrap confidence intervals (with 100 replications) are provided in square brackets.
Significant estimates at 5% significance level are denoted *.

for those individuals at the highest wage decile, the change in returns, had the distribution of regressors not changed, would have improved their wages. The signs of the contributions of both regressors and returns are in line with the results obtained in Melly (2005) for the U.S.

The counterfactual decomposition of the differences between the IQR shows that the observed increase in the inequality, mainly at the top of the distribution, has been driven by the changes in the returns. We find however that the observed changes in the distribution of the regressors from 1983 to 1993 are not significant in explaining the evolution of wage inequality between both periods. The percentage of overall growth of wage inequality than can be ascribed to changes in the regressors with our results is smaller than that found in Melly (2005); however, it is similar in magnitude to the percentage found in Juhn et al. (1993) who, by assuming independence between the errors and the regressors, do not account for heteroskedasticity of the errors as Melly (2005) and we do.

We have also performed the counterfactual analysis of this subsection using $\widehat{Q}_\tau^{t_X=r,t_\theta=r^*}(W)$ instead of $\widehat{Q}_\tau^{t_X=r^*,t_\theta=r}(W)$ on the right-hand side of (3.3), so that the returns are fixed at 1983 in the composition effect. Although similar qualitative results are obtained in both cases, this is not necessarily ensured because the differences in returns between the two periods could have been such that composition effect could have been different in both cases. The results obtained under exogeneity of schooling are very close to the results presented under endogeneity in Table 4 for each term. The reason for this similarity is the same as for the close results of the marginal wage quantile in both exogenous and endogenous models explained above.

5.2.2 Decomposition of differences keeping some regressors fixed

In our context, it is interesting to isolate the effects of schooling and the unobserved ability component from the total effect of the joint distribution of regressors on wages while returns are held fixed. We analyse both the effect of the observed change in the marginal distribution of schooling and the conditional distribution of schooling given the rest of exogenous individual attributes.

Following the notation in Section 3.2, let us consider the estimate of the τth counterfactual quantile of wages if only the marginal distribution of schooling is distributed as in 1983 and the rest of attributes and returns are fixed at 1993, which is denoted by $\widehat{Q}_\tau^{t_S=83,t_{(Z_1',V)|S}=93,t_\theta=93}(W)$, and the τth counterfactual quantile of wages if both schooling and the unobserved ability are distributed as in period 1983 and the rest of attributes and returns are fixed at period 1993, which is denoted by $\widehat{Q}_\tau^{t_{(S,V)}=83,t_{Z_1|(S,V)}=93,t_\theta=93}(W)$. Both counterfactual quantiles are estimated using the Epanechnikov kernel. The bandwidths for the Nadaraya-Watson weights are 1.5 for S

and 2.5 for V; these bandwidths guarantee sufficient observations around each data point. These two counterfactual quantiles allow us to decompose the composition effect $[\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)]$ into three different terms: $[\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_S=83,t_{(Z_1',V)|S}=93,t_\theta=93}(W)]$, which is the part of the difference that is explained by changes in the marginal distribution of schooling, $[\widehat{Q}_\tau^{t_S=83,t_{(Z_1',V)|S}=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_{(S,V)}=83,t_{Z_1|(S,V)}=93,t_\theta=93}(W)]$, which is the part of the difference that is explained by changes in the conditional distribution of the unobserved component affecting schooling V, given S; and $[\widehat{Q}_\tau^{t_{(S,V)}=83,t_{Z_1|(S,V)}=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)]$, which is the part of the difference that is explained by changes in the conditional distribution of the control variables Z_1, given S and V. In the first term, we compare the τth wage quantile of two populations that only differ in the proportion of individuals at each educational group. Given that the conditional distribution of (Z_1, V) given S is the same in both populations, a different distribution of schooling arises due to a change in the conditional distribution of education S given (Z_1, V). Then, either changes in the distribution of Z_2 given (Z_1, V) or changes in the parameters of the schooling decision process in (2.5), or both, can drive the observed differences in the conditional distribution of education given (Z_1, V).

Table 5a reports the results corresponding to this decomposition of the composition effect, where each column corresponds to each of the terms above. In the second column, we observe that the counterfactual distribution of wages with the marginal distribution of schooling as in 1983 and the rest of the individual attributes and returns as in 1993 is dominated by the marginal distribution of wages in 1993. The change in the marginal distribution of schooling from 1983 to 1993 towards a more educated population, holding the returns and the distribution of the rest of individual attributes fixed, increased hourly wages around 1.8% at the bottom quantiles and by 3.3% at the top quantiles. These changes account for the 24% (for the lowest quantiles) and the 58% (for the top quantiles) of the overall composition effect.

Although the change in the joint distribution of regressors does not significantly explain the increase in wage inequality, by decomposing the role of each regressor, we find that the change in the marginal distribution of schooling is the only factor that significantly explains the increase in wage inequality (note the positive impact on $IQR_{0.9,0.1}$, denoting that the difference between the top decile and the bottom decile is 1.5% higher when schooling is marginally distributed as in 1993); this effect mainly arises because the marginal change in schooling would have increased inequality below the median. These results are in line with the conventional belief of the wage inequality increasing effect of education (see Buchinsky 1994 or Machado and Mata 2005, among others). The interpretation of the unconditional effect on wages of changes in the distribution of a covariate needs some care, because both the returns over the wage distribution and how the

Table 5a: Decomposition of the changes in the wage distribution that can be attributed to changes in the joint distribution of regressors (endogenous model)

Quantile τ	Change in Regressors: $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W)-\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$	Change in Marginal of Schooling: $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W)-\widehat{Q}_\tau^{t_S=83,t_{(Z_1',V)\mid S}=93,t_\theta=93}(W)$	Change in Distribution of $V\mid S$: $\widehat{Q}_\tau^{t_S=83,t_{(Z_1',V)\mid S}=93,t_\theta=93}(W)-\widehat{Q}_\tau^{t_{(S,V_\alpha)}=83,t_{Z_1\mid(S,V)}=93,t_\theta=93}(W)$	Change in Distribution of $Z_1\mid(S,V)$: $\widehat{Q}_\tau^{t_{(S,V)}=83,t_{Z_1\mid(S,V)}=93,t_\theta=93}(W)-\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$
10th quantile	0.0774* [0.0263;0.1285]	0.0189* [0.0090;0.0288]	-0.0154* [-0.0276;-0.0032]	0.0739* [0.0237;0.1240]
30th quantile	0.0809* [0.0447;0.1171]	0.0257* [0.0158;0.0356]	-0.0167* [-0.0268;-0.0065]	0.0719* [0.0376;0.0376]
Median	0.0751* [0.0432;0.1071]	0.0296* [0.0197;0.0395]	-0.0146* [-0.0247;-0.0045]	0.0601* [0.0301;0.0901]
70th Quantile	0.0745* [0.0432;0.1059]	0.0334* [0.0235;0.0433]	-0.0116* [-0.0224;-0.0008]	0.0527* [0.0229;0.0825]
90th Quantile	0.0568* [0.0156;0.0981]	0.0334* [0.0235;0.0433]	-0.0070 [-0.0180;0.0041]	0.0304 [-0.0112;0.0719]
Quantiles τ,τ'	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W)-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W)-\widehat{IQR}_{\tau,\tau'}^{t_S=83,t_{(Z_1',V)\mid S}=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_S=83,t_{(Z_1',V)\mid S}=93}(W)-\widehat{IQR}_{\tau,\tau'}^{t_{(S,V)}=83,t_{Z_1\mid(S,V)}=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_{(S,V)}=83,t_{Z_1\mid(S,V)}=93}(W)-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$
$\tau=0.9,\tau'=0.1$	-0.0206 [-0.0768;0.0356]	0.0145* [0.0044;0.0246]	0.0084 [-0.0016;0.0184]	-0.0435 [-0.0994;0.0125]
$\tau=0.5,\tau'=0.1$	-0.0023 [-0.0418;0.0372]	0.0107* [0.0035;0.0179]	0.0008 [-0.0075;0.0083]	-0.0138 [-0.0533;0.0257]
$\tau=0.9,\tau'=0.5$	-0.0183 [-0.0522;0.0156]	0.0038 [-0.0036;0.0111]	0.0076* [0.0029;0.0123]	-0.0297 [-0.0627;0.0034]

Note: 95% bootstrap confidence intervals (with 100 replications) are provided in square brackets.
Significant estimates at 5% significance level are denoted *.

Table 5b: Decomposition of the changes in the wage distribution that can be attributed to changes in the joint distribution of regressors (exogenous model)

Quantile τ	Exogenous Model		
	Change in Regressors: $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$	Change in Marginal of Schooling: $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_S=83,t_{Z_1'|S}=93,t_\theta=93}(W)$	Change in Distribution of $Z_1|S$: $\widehat{Q}_\tau^{t_S=83,t_{Z_1'|S}=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$
10th quantile	0.0833* [0.0399;0.1267]	0.0172* [0.0034;0.0311]	0.0661* [0.0266;0.1055]
30th quantile	0.0819* [0.0511;0.1128]	0.0243* [0.0106;0.0381]	0.0576* [0.0315;0.0838]
Median	0.0723* [0.0447;0.0999]	0.0297* [0.0155;0.0439]	0.0426* [0.0204;0.0649]
70th Quantile	0.0680* [0.0400;0.0959]	0.0323* [0.0180;0.0466]	0.0357* [0.0125;0.0589]
90th Quantile	0.0573* [0.0234;0.0912]	0.0299* [0.0176;0.0421]	0.0274 [-0.0040;0.0589]
Quantiles τ,τ'	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_S=83,t_{Z_1'|S}=93,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_S=83,t_{Z_1'|S}=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$
$\tau=0.9,\tau'=0.1$	-0.0260 [-0.0714;0.0194]	0.0126* [0.0029;0.0224]	-0.0386 [-0.0817;0.0045]
$\tau=0.5,\tau'=0.1$	-0.0110 [-0.0433;0.0213]	0.0124* [0.0051;0.0198]	-0.0234 [-0.0554;0.0085]
$\tau=0.9,\tau'=0.5$	-0.0150 [-0.0401;0.0101]	0.0002 [-0.0070;0.0074]	-0.0152 [-0.0383;0.0079]

Note: 95% bootstrap confidence intervals (with 100 replications) are provided in square brackets.
Significant estimates at 5% significance level are denoted *.

change in the distribution of the regressor over its support must be taken into account. Thus, although the estimated returns of schooling for 1993 are higher for the bottom wage quantile than for the top, the change in education over both periods was more concentrated at the top of the wage distribution. In fact, the average increase in years of education is 0.73 for those individuals at the 7th and 8th deciles of 1993 wages and 0.60 for those individuals in the 5th and 6th decile. This contrasts with the decrease of 0.40 for those individuals in the 3rd and 4th decile. This explains why the change in schooling increases wage inequality if returns are assumed to be fixed at 1993. The impact on inequality of the change in schooling and the change in the conditional distribution of the rest of the regressors are of opposite signs.

The results reported in the third column of Table 5a imply that individuals, mainly at the top, would earn more if both schooling and ability were distributed as in 1983 than if only the marginal distribution of schooling was as in 1983. The change in the distribution of V given S goes in the opposite direction of schooling and it compensates almost completely the effect of the change in the marginal distribution of schooling for the individuals at the bottom of the distribution, it compensates only around half of the effect for the individuals at the middle of the distribution, and has no effect for the individuals at the top of the distribution. Thus, considering the results in the second and third columns jointly, the evolution of the joint distribution of schooling and ability between years 1983 and 1993 did not have any effect on wages at the bottom of the distribution, although it significantly increased wages by 2.7% for those individuals at the top.

For comparative purposes, Table 5b reports the results that are found when a similar decomposition is performed, but assuming that education is exogenous. In this case it is not possible to analyze the part of the difference that can be attributed to changes in the distribution of the unobserved ability V, since this variable is not included in the exogenous model; note that, if our endogenous model is correctly specified, then in Table 5b the part of the changes in the wage distribution that can be attributed to changes in the unobserved ability is wrongly attributed to changes in the distribution of Z_1 given schooling. The results in Table 5a allow for a deeper understanding of the changes in the wage distribution than those in Table 5b. In fact, these results show that the joint evolution of the distribution of schooling and unobserved ability is crucial to understand the contribution of each regressor to explain the changes in the wage distribution during the eighties in the US. Once we take into account the joint evolution of schooling and unobserved ability, we find that the change in the bivariate distribution of S and V did not have any impact at the bottom quantiles of wages, and a much smaller impact at the top quantiles than the effect of the change in the marginal distribution of schooling.

Section 3 only discusses the case when the set of explanatory variables is $X = (S, Z_1', V)'$. However, all the definitions that are given in Section 3 can also be applied, *mutatis mutandis*, if we consider $X = (S, Z')'$ (note that assumptions (2.1)-(2.3) allow us to identify $Q_\tau(W \mid S, Z)$). Specifically, if we take $X_1 = Z$ and $X_2 = S$, we can consider the τth counterfactual quantile of wages if the conditional distribution of schooling, given the attributes Z, is distributed as in period 1983, but the returns and the distribution of Z are fixed in 1993; it is denoted as $Q_\tau^{t_{S|Z}=83, t_Z=93, t_\theta=93}(W)$. This counterfactual quantile is of interest in order to isolate the effect on wages of a change in the conditional distribution of education given the individual attributes in Z. In principle, the estimation of this counterfactual quantile could be performed as described in subsection 3.2. However, given the large dimension of Z, the Nadaraya-Watson estimate $\widehat{F}_{S|Z=z}^{t=83}(s)$ is affected by the curse of dimensionality. In order to circumvent this problem, we introduce the following additional assumption: for every $\alpha \in (0, 1)$, there exist parameters $\mu_\alpha \in \mathbb{R}$ and $\pi_\alpha \in \mathbb{R}^p$ such that $Q_\alpha^{t=83}(S \mid Z = z) = \mu_\alpha^{t=83} + z'\pi_\alpha^{t=83}$. With this additional assumption, we can ensure that $\widetilde{F}_{S|Z=z}^{t=83}(s) \equiv M^{-1} \sum_{j=1}^{M} 1\{\widehat{\mu}_{\alpha(j)}^{t=83} + (\widehat{\pi}_{\alpha(j)}^{t=83})'z \leq s\}$ is a consistent estimate of $F_{S|Z=z}^{t=83}(s)$, where $\alpha(j) \equiv j/(M + 1)$, M is a large enough number (e.g. $M = N_{83}$), and $\widehat{\mu}_{\alpha(j)}^{t=83}$, $\widehat{\pi}_{\alpha(j)}^{t=83}$ are consistent estimates of $\mu_{\alpha(j)}^{t=83}$, $\pi_{\alpha(j)}^{t=83}$. The estimator of the counterfactual quantile $Q_\tau^{t_{S|Z}=83, t_Z=93, t_\theta=93}(W)$ using $\widetilde{F}_{S|Z=z}^{t=83}(s)$ instead of $\widehat{F}_{S|Z=z}^{t=83}(s)$ is performed as follows: i) obtain a draw of size m from the uniform distribution on $(0, 1)$, denoted as $\{\xi_i\}_{i=1}^{m}$; ii) obtain a draw of size m with replacement from $\{Z_i^{t=93}\}_{i=1}^{N_{93}}$, denoted as $\{Z_{*i}^{t=93}\}_{i=1}^{m}$; iii) for $i \in \{1, ..., m\}$, obtain a draw from $\{\widehat{\mu}_{\alpha(j)}^{t=83} + (\widehat{\pi}_{\alpha(j)}^{t=83})'Z_{*i}^{t=93}\}_{j=1}^{M}$; this draw is denoted as $S_{*i}^{t_{S|Z}=83}$; iv) find the τth sample quantile of $\{\widehat{Q}_{\xi_i}^{t=93}(W \mid S = S_{*i}^{t_{S|Z}=83}, Z = Z_{*i}^{t=93}\}_{i=1}^{m}$. This estimator is denoted as $\widehat{Q}_\tau^{t_{S|Z}=83, t_Z=93, t_\theta=93}(W)$.

We could further decompose the joint distribution of Z: let us assume that we are interested in the distribution of wages that would have prevailed if workers had been paid according to the wage schedule observed in period 1993, and both S and Z_2 (conditional on Z_1) are distributed as in period 1983, but the rest of the individual attributes in Z_1 remain distributed as in 1993. The corresponding τth quantile of this distribution is denoted by $Q_\tau^{t_{S|(Z_1', Z_2)}=83, t_{Z_2|Z_1}=83, t_{Z_1}=93, t_\theta=93}(W)$. We also aim to estimate this counterfactual quantile by assuming a parametric specification for the distribution of Z_2 given $Z_1 = z_1$. In our application, the excluded variables Z_2 contain the variable distance to college *proximity* and its interactions with a subset of variables in Z_1; in this case, assuming a distribution for Z_2 given $Z_1 = z_1$ amounts to specifying a particular parametric distribution for the binary variable *proximity* given $Z_1 = z_1$. This conditional distribution is estimated using a probit model.

These counterfactual quantiles allow us to decompose $[\widehat{Q}_\tau^{t_X=93, t_\theta=93}(W) - \widehat{Q}_\tau^{t_X=83, t_\theta=93}(W)]$

into three different terms: $[\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_{S|Z}=83,t_Z=93,t_\theta=93}(W)]$, which is the part of the difference that is explained by changes in conditional distribution of schooling given the observed individual characteristics in Z; $[\widehat{Q}_\tau^{t_{S|Z}=83,t_Z=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_{S|(Z_1',Z_2)}=83,t_{Z_2|Z_1}=83,t_{Z_1}=93,t_\theta=93}(W)]$, which is the part of the difference that is explained by changes in the conditional distribution of the distance to college given Z_1; and $[\widehat{Q}_\tau^{t_{S|(Z_1',Z_2)}=83,t_{Z_2|Z_1}=83,t_{Z_1}=93,t_\theta=93}(W) - \widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)]$, which is the part of the difference that is explained by changes in the marginal distribution of the control variables Z_1. Table 6 reports this decomposition of the composition effect, where each column corresponds to each of the terms above.

In the counterfactual exercise performed in the first term, under the assumption that both the returns to individual characteristics Z and the CDF of Z are held fixed, a change in the conditional distribution of schooling given Z can arise in our model from either a change in the conditional distribution of unobserved ability V given Z or a change in the parameters $(\mu, \pi')'$ in the schooling decision model described in (2.5) between both years, or both. The results in the second column show that this change increased hourly wages not monotonically in wage quantiles: around 6.7% for the bottom wage quantile, around 5% for the median wage and almost 9% for the top wage quantiles. This explains the increase in wage inequality at the top of the wage distribution.

The impact on wages of the change in the distribution of distance to college given the individual attributes in Z_1, while holding fixed the CDF of Z_1 and returns in 1993 and the conditional distribution of schooling in 1983, can be found in the third column in Table 6. These results show that the effect of the observed changes in the distribution of distance to college, given the rest of schooling determinants, does not have any statistical significant impact on the distribution of wages. Note that one possible reason for the change in the distribution of *proximity* given Z_1 is the increase in the number of colleges during this period. The results in this table suggest that the changes in wage inequality are mainly explained by the changes in the distribution of the unobserved ability or by changes in the process that determines individual educational attainment, and not by changes in the distribution of distance to college.

Comparing the role of schooling in Tables 5 and 6 shows that the effect of a change in the distribution of education conditional on its observable determinants has higher impact on all the quantiles than a change in the marginal distribution of schooling; they also explain different parts of the increase in wage inequality. The estimates in both tables capture the effect of the observed changes in the schooling decision process during this time period. The change in the marginal distribution of schooling is additionally capturing the effect of changes in the conditional distribution of distance to college given (Z_1', V); the effect of the change

Table 6: Decomposition of the changes in the wage distribution that can be attributed to changes in the joint distribution of regressors

Quantile τ	Change in Regressors: $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$	Change in Conditional of Schooling given Z_1, Z_2: $\widehat{Q}_\tau^{t_X=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_{S\mid(Z_1',Z_2)}=83,t_{(Z_1',Z_2)}=93,t_\theta=93}(W)$	Change in Distribution of $Z_2\mid Z_1$: $\widehat{Q}_\tau^{t_{S\mid(Z_1',Z_2')}=83,t_{(Z_1',Z_2)}=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_{(S,Z_2)\mid Z_1}=83,t_{Z_1}=93,t_\theta=93}(W)$	Change in Distribution of Z_1: $\widehat{Q}_\tau^{t_{(S,Z_2)\mid Z_1}=83,t_{Z_1}=93,t_\theta=93}(W)-$ $-\widehat{Q}_\tau^{t_X=83,t_\theta=93}(W)$
10th quantile	0.0774* [0.0263;0.1285]	0.0675* [0.0346;0.1005]	-0.0141 [-0.0529;0.0247]	0.0240 [-0.0262;0.0741]
30th quantile	0.0809* [0.0447;0.1171]	0.0352* [0.0088;0.0616]	-0.0064 [-0.0461;0.0333]	0.0521* [0.0221;0.0820]
Median	0.0751* [0.0432;0.1071]	0.0485* [0.0201;0.0769]	-0.0092 [-0.0295;0.0111]	0.0358* [0.0077;0.0638]
70th Quantile	0.0745* [0.0432;0.1059]	0.0731* [0.0400;0.1063]	-0.0174 [-0.0393;0.0045]	0.0187 [-0.0115;0.0488]
90th Quantile	0.0568* [0.0156;0.0981]	0.0890* [0.0497;0.1282]	-0.0151 [-0.0409;0.0107]	-0.0171 [-0.0592;0.0250]
Quantiles τ,τ'	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_{S\mid(Z_1',Z_2)}=83,t_{(Z_1',Z_2)}=93,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_{S\mid(Z_1',Z_2')}=83,t_{(Z_1',Z_2)}=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_{(S,Z_2)\mid Z_1}=83,t_{Z_1}=93,t_\theta=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_{(S,Z_2)\mid Z_1}=83,t_{Z_1}=93,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)$
$\tau=0.9, \tau'=0.1$	-0.0206 [-0.0768;0.0356]	0.0214 [-0.0161;0.0590]	-0.0010 [-0.0378;0.0358]	-0.0410 [-0.0837;0.0017]
$\tau=0.5, \tau'=0.1$	-0.0023 [-0.0418;0.0372]	-0.0190 [-0.0486;0.0106]	0.0049 [-0.0295;0.0394]	0.0118 [-0.0436;0.0672]
$\tau=0.9, \tau'=0.5$	-0.0183 [-0.0522;0.0156]	0.0404* [0.0136;0.0673]	-0.0059 [-0.0384;0.0266]	-0.0529 [-0.0848;-0.0209]

Note: 95% bootstrap confidence intervals (with 100 replications) are provided in square brackets.
Significant estimates at 5% significance level are denoted *.

in the conditional distribution of schooling given Z is also estimating the effect of changes in the conditional distribution of unobserved innate ability V given Z. Therefore, the different results in both decompositions are explained by the higher impact on wages of changes in the conditional distribution of unobserved ability than the impact on wages of changes in the conditional distribution of distance to college.

5.2.3 Decomposition of differences keeping some returns fixed

Table 7 reports the results that are found when the changes in the wage distribution that can be attributed to changes in the returns are decomposed as described in (3.6). First of all, note that none of the terms in the decomposition of the differences between quantiles is statistically significant. It is not surprising that the differences attributed to changes in the returns to schooling (second column) are not significant, since in Figure 1 we obtain that the differences between the estimates of the returns to education are not significantly different from zero. The point estimates imply that the change in the returns to schooling from 1983 to 1993 increased wages at the bottom of the distribution and decreased wages at the top of the distribution. This is in line with the changes in the returns to schooling obtained for the endogenous model shown in Figure 1. However, this implies a statistically significant decrease in the wage inequality due to the change in the returns to schooling, as the bottom part in the second column show. The changes in the returns to ability had the opposite effect: a negative impact on the bottom part of the wage distribution and a positive impact at the upper part. The increase in wage inequality attributed to changes in returns is explained by the change in the returns to Z_1, which increased inequality slightly more than the changes in the returns to schooling decreased inequality. We also find that the changes in the returns to ability are not statistically significant in order to explain the observed changes in wage inequality.

The results of separating the effects of coefficients from the effects of residuals suggested by Melly (2005, p.581) show that the changes in the within-group inequality are the only significant factor explaining the significant changes in wage inequality that arise from changes in the returns; also, the decrease in wages at the bottom and median quantiles if only the returns to skills would have changed from 1983 to 1993 holding the distribution of regressors fixed is explained by changes in the median coefficients. It is important to stress that the evolution of the distribution of the residuals account for a larger percentage of the overall change in inequality than in Melly (2005): residuals account for around 90% of that part of the change in wage inequality explained by changes in returns. Although we also correct for the composition effect, our results show that most of the wage inequality arises from a change in the distribution of the residuals as is standard

Table 7: Decomposition of the changes in the wage distribution that can be attributed to changes in the returns (decomposition return by return; endogenous model)

Quantile τ	Change in Returns: $\widehat{Q}_{\tau}^{t_X=83,t_\theta=93}(W)-$ $-\widehat{Q}_{\tau}^{t_X=83,t_\theta=83}(W)$	Change in Returns to Schooling: $\widehat{Q}_{\tau}^{t_X=83,t_\theta=93}(W)-$ $\widehat{Q}_{\tau}^{t_X=83,t_\beta=83,t_{(\gamma,\lambda)}=93}(W)$	Change in Returns to unobservables V: $\widehat{Q}_{\tau}^{t_X=83,t_\beta=83,t_{(\gamma,\lambda)}=93}(W)-$ $\widehat{Q}_{\tau}^{t_X=83,t_{(\beta,\lambda)}=83,t_\gamma=93}(W)$	Change in Returns to Z_1: $\widehat{Q}_{\tau}^{t_X=83,t_{(\beta,\lambda)}=83,t_\gamma=93}(W)-$ $-\widehat{Q}_{\tau}^{t_X=83,t_\theta=83}(W)$
10th quantile	-0.0765* [-0.1145;-0.0385]	0.9019 [-0.8107;2.6145]	−0.0044 [-0.0648;0.0559]	−0.9740 [-2.6953;0.7473]
30th quantile	-0.0544* [-0.0774;-0.0314]	0.4171 [-0.8413;1.6756]	0.0032 [-0.0439;0.0502]	−0.4746 [-1.7563;0.8071]
Median	-0.0451* [-0.0687;-0.0215]	0.1410 [-1.0090;1.2911]	−0.0063 [-0.0462;0.0335]	−0.1798 [-1.3315;0.9719]
70th Quantile	-0.0192 [-0.0429;0.0045]	−0.5793 [-1.8514;0.6929]	0.0012 [-0.0457;0.0481]	0.5589 [-0.7036;1.8213]
90th Quantile	0.0722* [0.0373;0.1071]	−1.2424 [-2.9302;0.4454]	0.0742 [-0.0455;0.1939]	1.2404 [-0.3936;2.8743]
Quantiles τ,τ'	$\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=83}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=93}(W)-$ $\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\beta=83,t_{(\gamma,\lambda)}=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\beta=83,t_{(\gamma,\lambda)}=93}(W)-$ $\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_{(\beta,\lambda)}=83,t_\gamma=93}(W)$	$\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_{(\beta,\lambda)}=83,t_\gamma=93}(W)-$ $-\widehat{IQR}_{\tau,\tau'}^{t_X=83,t_\theta=83}(W)$
$\tau=0.9,\tau'=0.1$	0.1487* [0.0527;0.2446]	−2.1443∗ [-3.2292;-1.0594]	0.0786 [-0.0434;0.2007]	2.2143∗ [1.1426;3.2861]
$\tau=0.5,\tau'=0.1$	0.0314 [-0.0363;0.0990]	−0.7609∗ [-1.4108;-0.1110]	−0.0019 [-0.0577;0.0539]	0.7942∗ [0.1679;1.4205]
$\tau=0.9,\tau'=0.5$	0.1173* [0.0571;0.1775]	−1.3834∗ [-2.1780;-0.5889]	0.0806 [-0.0316;0.1927]	1.4202∗ [0.6613;2.1791]

Note: 95% bootstrap confidence intervals (with 100 replications) are provided in square brackets.
Significant estimates at 5% significance level are denoted *.

in the literature (see Juhn et al. 1993).

6 Concluding remarks

In this paper we have performed a counterfactual decomposition of the changes in wages by disentangling the effect of changes in the joint distribution of the individual characteristics and the effect explained by changes in the prices of these attributes (i.e. changes in the wage structure) by extending the works of Machado and Mata (2005) and Melly (2005) to the case where the schooling decisions are regarded as endogenous to earnings. Although controlling for endogeneity is common when estimating wage equations, it has not been previously taken into account in the construction of counterfactual decompositions. The goal is to obtain the marginal wage quantile consistent with the conditional quantile function model of wages given schooling, exogenous attributes and ability and define those counterfactual distributions when the joint, conditional and marginal distribution of the regressors are fixed for a comparison year or group.

This methodology is used to analyze the sources of the changes in wage distribution that took place in the United States between 1983 and 1993. In the estimation of the conditional quantile model, we obtain the quantile structural effects of schooling on wages under the control function assumption by using proximity to college for different parental education as instruments. We discuss the advantages of using the control function (CF) approach for our context over the instrumental variable (IV) approach: on the one hand, the IV assumption only allows for heteroskedasticity in Z_1 while the CF assumption does not exclude heteroskedasticity of the errors depending on V, S and Z_1; on the other hand, the CF assumption allows us to identify the conditional quantile function of wages given schooling and other individual attributes in order to obtain the counterfactual distributions.

Our decomposition results show that the observed change in the joint distribution of the regressors would have increased wages at all quantiles without any significant impact on wage inequality. However, if only returns to skills had changed and the distribution of the attributes had remained as in 1983, all but the top quantile would have decreased and this change would have had a significant and positive effect on the increase in wage inequality.

As opposed to the results in Melly (2005), we obtain a small contribution of the compositional effects in explaining the increase in wage inequality, which is not statistically significant, despite the fact that we allow for the whole distribution of residuals to depend on regressors. This is due to the fact that while the change in the joint distribution of schooling and ability experienced during the eighties would have significantly increased wage inequality, the effect of the change in the joint distribution of the rest of the regressors goes in the opposite direction. Schooling

on its own has provided to be a factor that lead to a more unequal wage distribution. As in the previous literature, the change in the distribution of the conditional distribution of the residuals is the main factor explaining the increase in the wage inequality while the changes in the median returns are not important to explain the observed changes in the wage distribution.

Our results suggest that schooling is endogenous, since the quantiles of the unobservables affecting wages are a function of the unobserved ability affecting education. In any case, for comparison reasons, we perform the decomposition of the changes in wages under both the assumption of endogeneity and exogeneity of education. Some differences between the results from the endogenous and the exogenous model arises in the decomposition of the part of the overall change due to changes in returns into those parts explained by changes in the coefficients of each subset of individual attributes. The coefficients of education can be interpreted as the returns to schooling only under the control function approach. Once we control for the endogeneity of schooling, we observe a decrease in the returns of schooling at the top of the distribution and an increase at the bottom of the distribution during the decade under consideration. Both effects consequently imply our result that, if only the returns to schooling had changed between both periods, wage inequality would have decreased. A larger and opposite effect is found for the changes in the returns of the rest of the exogenous variables, so that the overall impact of the changes in returns is positive.

We also provide alternative decompositions of the aggregate effect of regressors on wages into the role of the change in the distribution of each covariate of interest. Although the aggregate contribution of the regressors and coefficients are similar in both the endogenous and the exogenous model, accounting for the endogeneity of schooling allows us to perform those counterfactual distributions that would have prevailed if ability given schooling or if schooling given ability had been distributed as in the comparison year. It is not possible to obtain these distributions under the exogenous model, where only the marginal distribution of schooling can be fixed.

The analysis of the impact of the change in the conditional distribution of ability given schooling shows that its negative effect on wages compensates the positive effect at the lower wage quantile of the change in the marginal distribution of schooling. Therefore, the change in the joint distribution of wages and ability did not have any effect at the bottom part of the wage distribution; it has a positive effect at the top decile driven by an increase of 3.3% due to the change in the marginal distribution of schooling, while the change in the distribution of ability given education did not have any effect. The effect of schooling at the top quantile is reduced to 1.45% once we condition on ability. The part of the changes in wages explained by

the changes in ability is misleadingly attributed to the rest of the regressors if the exogenous model is estimated.

When the aggregate effect of regressors is decomposed into the effect of schooling, given the rest of individual attributes, and the effect of the joint distribution of these regressors, we find that the changes in the conditional distribution of unobserved ability or the changes in the parameters of the model leading the individual schooling decision are more important in explaining the increase in wage inequality than the changes in the distribution of individual attributes. Comparing the effect of the change in the marginal distribution of schooling and the effect of the change of the conditional distribution of education, given its determinants, allows us to conclude that the change in the conditional distribution of unobserved ability produces a higher shift in the wage distribution than the change in the conditional distribution of distance to college, given individual attributes and ability. However, this latter change in the distribution of ability explains the overall increase in wage inequality better.

APPENDIX: Definition of Variables

The measure of wage W used is the logarithm of the hourly wage, which is computed as the ratio between the total annual amount of labor income (total wages and salaries, including tips and commissions) from the previous year and the total annual hours that the individual reports to have worked for money in the previous year. For comparative purposes, we deflate wages to 1982-1984 dollars using the annual CPI-U (All Urban Consumers) Price Index.

The schooling variable S is defined as the number of school grades completed by the individual, which coincides with the number of years of schooling; it takes on 18 different values: from 0 if the individual achieved no grade to 17 (censoring point) if the individual finished some postgraduate activity (thus, S is 16 for college graduates, and less than 12 for high school dropouts).

The vector of exogenous variables Z_1 includes: tenure (self-reported number of years worked since the individual was 18), squared tenure, a race dummy (1 if the individual reports to be white), a dummy for the size of the city (1 if the size of the largest city in the current county of residence is larger than 50,000 inhabitants), a dummy for government occupation (1 if the individual reports to work for the federal, state or local government), 3 regional dummies (for south, north and northeast regions; west is the reference region), 11 industry dummies (public administration industry is used as the reference group) and 14 dummies, F_j, M_j, for $j = 2, ..., 8$, to capture parental education, where F_j (or M_j) is a dummy variable which takes value 1 only if the father (or mother) of the head has completed at most education level j, where education level 2 is from the 6th to the 8th grade, education level 3 is from the 9th to the 11th grade, education level 4 is high school, education level 5 is high school plus nonacademic training, education level 6 is some college (but not completed) or an associate degree, education level 7 is a college BA and education level 8 is finished college, advanced or professional degree or some graduate work (note that the reference group are parents who only completed the 5th grade at most).

The vector of excluded instruments Z_2 includes dummy distance to college *proximity* as defined in Section 4 and its interaction with parental education dummies, i.e. *proximity* $* F_j$ and *proximity* $* M_j$, for $j = 2, ..., 8$.

REFERENCES

Albrecht, J., A. Bjorklund and S. Vroman (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics 21*, 145-177.

Albrecht, J., A. van Vuuren and S. Vroman (2009). Counterfactual distributions with sample selection adjustments: Econometric Theory and an application to the Netherlands. *Labor Economics 16*, 383-396.

Angrist J.D., V. Chernozhukov and I. Fernández-Val (2006). Quantile regression under misspecification with an application to the US Structure Wage. *Econometrica 74*, 539-563.

Angrist, J.D. and A.B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *Quaterly Journal of Economics 106*, 979-1014.

Arulampalam, W., A.L. Booth and M.L. Bryan (2007). Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review 60,* 163-186.

Autor, D.H., L.F. Katz and M.S. Kearney (2005). Rising wage inequality: the role of composition and prices. NBER Working Paper 11628.

Blundell, R. and Powell, J. (2003), Endogeneity in Nonparametric and Semiparametric Regression Models. Chapter 8. Advances in Economics and Econometrics, M. Dewatripont, Hansen, L. and S. J. Turnsovsky (eds.), Cambridge University Press.

Brunello, G., M. Fort and G. Weber (2009). Changes in compulsory schooling, education and the distribution of wages in Europe. *Economic Journal 119,* 516-539.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: application of quantile regression. *Econometrica 62,* 405-458.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Economics: Essays in Honour of John Vanderkamp,* ed. by L. Christofides, E.K. Grant and R. Swindinsky. University of Toronto Press.

Carneiro, P. and J.J. Heckman (2002). The evidence on credit constraints in post-secondary schooling. *Economic Journal 112,* 705-734.

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*, 191-208.

Chaudhuri, P. (1991). Nonparametric estimation of regression quantiles and their local Bahadur representation. *Annals of Statistics 19*, 760-777.

Chernozhukov, V., I. Fernández-Val and B. Melly (2009). Inference on counterfactual distributions. Cemmap Working Papers CWP09/09.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects, *Econometrica* 73, 245-261.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models, *Journal of Econometrics* 132, 491-525.

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

DiNardo, J., N.M. Fortin and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. *Econometrica 64*, 1001-1044.

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

Firpo, S., N. Fortin and T. Lemieux (2007). Decomposing Wage Distributions using Recentered Influenced Function Regressions. Mimeo.

Gosling, A., S. Machin and C. Meghir (2000). The changing distribution of male wages in the UK. *Review of Economic Studies 67*, 635-666.

Gottschalk, P. and T.S. Smeeding (1997). Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature 35*, 633-687.

Honoré, B.E. and L. Hu (2004). On the performance of some robust instrumental variables estimators. *Journal of Business and Economic Statistics 22,* 30-39.

Juhn C., M.K. Murphy and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy 101*, 410-442.

Kling, J. (2001). Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business and Economic Statistics 19*, 358-364.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory 19*, 1-31.

Lee, S. (2004). Endogeneity in quantile regression models: a control function approach. *Cemmap Working paper* CWP08/04

Lee, S. (2007). Endogeneity in quantile regression models: a control function approach. *Journal of Econometrics 141,* 1131-1158.

Levy, F. and R.J. Murnane (1992). U.S. earnings levels and earnings inequality: a review of recent trends and Proposed Explanations. *Journal of Economic Literature 30*, 1333-1381.

Newey, W.K., J.L. Powell and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*, 565-603.

Ma, L. and R. Koenker (2006). Quantile regression methods for recursive structural equation models. *Journal of Econometrics 134*, 471-506.

Machado, J.A.F. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 445-465.

Machado, J.A.F. and J.M.C Santos Silva (2005). Quantiles for Counts. *Journal of the American Statistical Association*, 1226-1237.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics 12,* 577-590.

Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. University of St. Gallen Discussion Paper.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14,* 693-709.

Yu, K. and M.C. Jones (1993). Local linear quantile regression. *Journal of the American Statistical Association 93*, 228-237.