Public and private sector wage distributions controlling for endogenous sector choice

Blaise Melly*
Swiss Institute for International Economics and Applied Economic Research (SIAW), University of St. Gallen

Abstract:

We apply the instrumental quantile regression model of Chernozhukov and Hansen (2005b) to examine the wage structure in the public and private sector. Since the original estimator does not allow for interaction terms between the endogenous variable and the covariates, we propose two estimators that allow for this possibility. The results assuming exogenous sector choice give a negative mean public sector wage premium and show that the wage distribution is more compressed in the public sector. Correcting for endogenous sector choice reverses the findings concerning the mean premium but preserves the more compressed structure of the public sector earnings distribution.

Keywords: Wage Inequality, Quantile Regression, Instrumental Variables, Wage Differentials, Public and Private Sector.

JEL classification: C13, C14, C21, J31, J45.

*Address for correspondence:
Blaise Melly
Bodanstrasse 8, 9000 St. Gallen, Switzerland
Phone: 0041712242301
Fax: 0041712242298
blaise.melly@unisg.ch

1. Introduction

Public sector pay attracts public attention for two reasons. First, public sector labor markets are large and the size of the public sector wage bill has implications for both monetary and fiscal policy. Second, public sector labor markets are different and there are a number of reasons that earnings differential between the private and the public sector exists. The public sector is subject to political constraints and not profit constraints. Therefore the political system may have different objectives from those of the private sector. Issues of pay equity and fairness can survive in the political market place more than in the economic market place.

Given these differences in the wage setting procedures and the possible consequences on the labor market, many researchers have sought to ascertain whether an identical employee working in the same job in the public and in the private sector would earn the same or a different amount. Early research (Smith 1976, Gunderson 1979) has used least-squares regressions to compare the predicted wages in both sectors. More recently, two different directions of research have complemented these first results. One strand of research takes account of possible non-random selection by Heckman/Lee correction for selectivity, endogenous switching regression models or fixed effect panel models (Goddeeris 1988, Gyourko and Tracy 1988, Dustmann and Van Soest 1998). They mostly find a significant selection bias but their results widely differ from one study to the other, probably because of weak (or absence of real) instruments. A second strand of research has compared the extent of earnings dispersion for public and private sector employees (Poterba and Rueben 1995, Mueller 1998, Melly 2005a). The main conclusion from these studies is that the public sector compresses the distribution of earnings.

The objective of this paper is to bring these two strands of research together. Until recently, it was impossible to allow at the same time for endogeneity of the sector choice and for heterogeneous public sector gap at different points of the distribution. During the last years, different instrumental variable methods for estimating endogenous quantile treatment effect have been proposed[1]. After comparing the different models and estimators, we use the proposal of Chernozhukov and Hansen (2004b) to estimate the effects of the public sector status on the entire wage distribution controlling for endogenous sector choice. We use data from the German Socio-Economic Panel, which contain background information on parents' economic status and provides us with reasonable instruments. The results indicate that there is considerable heterogeneity in the effect of the public sector status and that controlling for endogenous sector choice is important.

This paper contributes to the literature both methodologically and substantially. Since the estimator of Chernozhukov and Hansen does not allow for the estimation of interaction terms without additional instruments, we propose two estimators that allow to estimate fully interacted models. Their asymptotic distributions are derived and a Monte-Carlo simulation compares their finite-sample properties. This paper also contributes to the literature on the public-private sector wage differential. To the best of our knowledge, it is the first study that both control for endogenous sector choice and analyze the public and private sector wage distributions. Rigorous inference procedures show that heterogeneous public sector gaps and endogenous sector choice are really present and important. Applying the new estimators, we find that returns to education are higher in the private sector and that the experience-wage function is more concave in the private sector.

Section 2 compares the different models that have been proposed recently to correct for endogeneity in the quantile regression model and presents into more details the estimator of Chernozhukov and Hansen (2004b and 2005b). Section 3 proposes two extensions allowing for regressors fully interacted with the endogenous variable and gives the results of Monte-

[1] Abadie, Angrist and Imbens (2002), Chernozhukov and Hansen (2005b),Chesher (2003), Ma and Koenker (2005), Lee (2004), Honore and Hu (2004), Hong and Tamer (2003), among others.

Carlo simulations. Section 4 shows how it is possible to recover the unconditional wage distribution if we have estimated the conditional wage distribution by quantile regression and how to use this result to decompose differences in the wage distribution. Section 5 describes the data set, inclusively the instruments, along with some descriptive statistics. Section 6 presents the empirical results and section 7 concludes.

2. Endogeneity in the quantile regression model

The basic quantile regression model specifies the conditional quantile as a linear function of covariates. Let *Y* be the dependent variable of interest and *X* be a vector of exogenous explanatory variables. It is assumed that:

$$Y = X'\beta(\theta) + \varepsilon \text{ and } Q_\theta\left(\varepsilon \middle| X{=}x_i\right) = 0,$$

where $Q_\theta\left(\varepsilon \middle| X{=}x_i\right)$ denotes the θ^{th} quantile of ε conditional on $X = x_i$. Koenker and Bassett (1978) propose to estimate the θ^{th} regression quantile by solving

$$\hat{\beta}(\theta) = \underset{\beta\in\mathbb{R}^K}{arg\ min} \sum_{i=1}^{N} \rho_\theta\left(y_i - x_i'\beta\right),$$

where ρ_θ is the *check function:* $\rho_\theta(z) = z\left(\theta - 1(z \le 0)\right)$ and $1(\cdot)$ is the usual indicator function. They show the $\sqrt{N}$ consistency and asymptotic normality of $\hat{\beta}(\theta)$. Buchinsky (1991) shows that this estimator has a GMM interpretation since $\beta(\theta)$ satisfies the following moment condition

$$E\left[\left(\theta - 1\left(Y < X'\beta(\theta)\right)\right)X'\right] = 0.$$

Increasing θ continuously from 0 to 1, we can trace the entire distribution of *Y* conditional on *X*. By replacing a monolithic model of conditional central tendency with a family of models for conditional quantiles, we are able to achieve considerably greater flexibility and a much more complete view of the effect of the covariates on the dependent variable, allowing them to influence location, scale and shape of the response distribution. For instance, the distributional consequences of minimum wages, training programs and education are of primary interest to policy makers. Unfortunately, in most cases, the treatment is self-selected or endogenous, making conventional quantile regression inappropriate.

Amemiya (1982) was the first to seriously consider quantile regression methods in the presence of endogenous regressors. He shows the consistency and asymptotic normality of a class of two-stage median regression estimators. Subsequent work of Powell (1983) and Chen and Portnoy (1996) extended this approach but maintained the focus primarily on the conditional median problem. The main motivation of these works was the robustness of the median regression. Chernozhukov and Hansen (2001) show that this approach is not consistent when the quantile treatment effect differs across quantiles which is precisely the main motivation for using quantile regression.

Other approaches have been considered. Abadie, Angrist and Imbens (2002) use the LATE framework of Angrist, Imbens and Rubin (1996) and consider a quantile treatment effect model for the subpopulation of compliers. Their estimator applies only to a restrictive case - binary treatment variable and a single binary instrument – and imposes a monotonicity condition which is not likely to be satisfied in our application.

Chesher (2003) develops a general nonlinear model which may be viewed as an extension of the recursive causal chain models discussed by Strotz and Wold (1969). He shows the nonparametric identification of the parameters of interest. Based on his results, Ma and Koenker (2005) propose two estimators but they assume a finite-dimensional parametric restriction and integrate over the nonparametric estimates. The identification strategy of Chesher (2003) requires the dependent variable, the endogenous variables and the instruments to be continuous. Chesher (2005) shows that some extensions for discrete variables are possible but excludes the binary endogenous variable case, which is exactly the situation encountered in the public-private sector application.

A more direct estimation strategy uses the exclusion restrictions (instruments) directly in the GMM framework. Suppose we have a structural relationship defined by

$$Y = D'\alpha(U) + X'\beta(U), \qquad U|X,Z \sim \text{Uniform}(0,1), \tag{1}$$

$$\theta \to D'\alpha(\theta) + X'\beta(\theta) \text{ is strictly increasing in } \theta, \tag{2}$$

$$D = \delta(X,Z,V). \tag{3}$$

In these equations,

Y is the scalar outcome of interest,

U is a scalar unobserved random variable,

D is a vector of endogenous variables determined by (3), where

X is a vector of exogenous control variables,

Z is a vector of instrumental variables, and

V is a vector of unobserved random variables possibly correlated with *U*.

(1) and (2) imply that

$$\Pr\left(Y \le D'\alpha(\theta) + X'\beta(\theta) \middle| X,Z\right) = \theta, \tag{4}$$

thus providing the moment conditions

$$E\left[\left(\theta - 1\left(Y \le D'\alpha(\theta) + X'\beta(\theta)\right)\right)(X',Z')'\right] = 0. \tag{5}$$

Assuming iid sampling, compactness on the support of variables and on the parameter space, and some full rank conditions assuring that the parameters are identified[2], we could estimate $\alpha(\theta)$ and $\beta(\theta)$ by traditional GMM. This strategy was used by Hong and Tamer (2003), Chen, Linton and Keilegom (2003) and Honore and Hu (2004) to construct estimators. Hong and Tamer (2003) also present a discussion of conditions under which this model is identified.

Abadie (1995) noted the computational difficulty in obtaining the solution to the optimization problem. The objective function is "million-modal" and has zero derivative almost everywhere, implying the need to perform a grid search over a subset of $\mathbb{R}^{\dim(\alpha)+\dim(\beta)}$, thus rendering the application of this estimator almost impossible in data sets typically found in microeconometrics[3].

The instrumental variable quantile regression (IQR) estimator proposed by Chernozhukov and Hansen (2004b) can be viewed as a computationally attractive method of approximately solving the moment condition (5). Their basic idea is simple. If we knew the true coefficients $\alpha(\theta)$, we could estimate $\beta(\theta)$ consistently by regressing $Y - D'\alpha(\theta)$ on *X* with traditional quantile regression. In reality, we don't know $\alpha(\theta)$ but we have instruments *Z*. Thus, we can try different values for $\alpha(\theta)$ and regress $Y - D'\alpha(\theta)$ on *X* and *Z*. If the model is identified, the true value of $\alpha(\theta)$ is the only one for which the coefficients on *Z* are zero. This reduces considerably the computation time since we only need to perform a grid search on $\dim(\alpha)$ which is frequently small[4].

In this paper we apply and extend the approach of Chernozhukov and Hansen for three reasons. First, it is the more general approach and some other interesting estimators cannot be applied because the data don't satisfy their assumptions (a binary endogenous variable and 5 instruments). Second, it is computationally tractable since it requires solving traditional quantile regression on a grid search over parameter sets of dimension $\dim(\alpha)$. In the public-private sector application, there is only a one-dimensional endogenous variable. Moreover, quantile regression can be solved very fast using recently developed algorithms (Portnoy and

---

[2] See for instance assumption R2 in Chernozhukov and Hansen (2005b). Basically, it requires that a density-weighted covariance matrix between *D* and *Z* is of full rank.

[3] The same remark can be made about the robust LIMIL estimator of Sakata (2001). Moreover, the main objective of his estimator is to robustify the traditional LIMIL estimator, not to estimate the effect of endogenous variables on the distribution of the potential outcome.

[4] Details about the procedure and the asymptotic distribution of the estimator are given in Appendix A.

Koenker 1997). Third, only Chernozhukov and Hansen have derived the properties of the instrumental quantile regression process and have proposed consistent testing procedures derived from it. This is important in the present applications since significant results cannot be obtained if we consider a single quantile but are found using the process.

3. Extensions of the IQR

One drawback of the IQR is that we need at least one instrument for each endogenous regressor *inclusively* interaction terms with covariates. For instance, if we want to estimate not only a different constant between the public and private sector but also different rates of return to *X*, we need $\dim(\beta)+1$ instruments, which can be very difficult if not impossible. Therefore two possible estimators which allow for interacted terms are proposed in the next two sections. The first uses the sample selection correction of Buchinsky (1998) to estimate the slope parameters and the Chernozhukov and Hansen's estimator to estimate the constants. The second uses the Chernozhukov and Hansen's estimator locally at different point of the distribution of *X* and integrate then the results using the minimum distance framework.

3.1. Combination of sample selection and IV quantile regression

Buchinsky (1998, 2001) proposes a sample selection procedure for quantile regression. His estimator can be considered as the quantile regression equivalent of Newey (1988) series estimator. The key assumption is the single index restriction on the error term. Its distribution is assumed to depend on the regressors *X* and the instruments *Z* only through an index function, which can be estimated in a first step. The bias term can then be approximated by a power series of the estimated index. The constant is estimated "at infinity" using an idea suggested by Heckman (1990) and Andrews and Schafgans (1996) for the estimation of the constant term in mean regression.

Now, if we consider the public private sector application again, D is an endogenous dummy variable that is equal to 0 if the person works in the private sector and 1 if she works in the public sector. The idea of the estimator proposed in this section is to consider a fully interacted instrumental variable model as a switching regression model:

$$Y_0 = \alpha_0(\theta) + X'\beta_0(\theta) + \varepsilon_0$$

$$Y_1 = \alpha_1(\theta) + X'\beta_1(\theta) + \varepsilon_1$$

$$Y = DY_1 + (1-D)Y_0 .$$

Therefore, considering the two wage equations separately, we can estimate them using the sample selection correction of Buchinsky (1998). The constant terms can theoretically be estimated "at infinity" as proposed by Buchinsky if there are some observations with $\Pr(D=1) \to 1$ and others with $\Pr(D=1) \to 0$. Such a method has the drawbacks that it requires very strong, large support conditions and that estimation that directly follows the identification strategy involves estimation on "thin sets" and thus a slow rate of convergence. Identification hinges on sufficient distribution mass in the tails of the index which must be unbounded. The estimation often rests on just a handful of observations surpassing the growing threshold which may be hard to distinguish from unreasonable outliers.

For all these reasons we propose to estimate only the slope coefficients with the sample selection procedure of Buchinsky and we obtain $\sqrt{N}$ consistent and asymptotically normally distributed estimates $\hat{\beta}_0(\theta)$ and $\hat{\beta}_1(\theta)$. In a second step, both constant terms are estimated using a slightly modified version of the instrumental quantile regression estimator. We use the Chernozhukov and Hansen (2004b) estimator with $Y-(1-D)X\hat{\beta}_0 - DX\hat{\beta}_1$ as dependent variable and only a constant as exogenous regressor. Using traditional results for sequential GMM estimators we can prove that $\hat{\beta}_0(\theta)$, $\hat{\beta}_1(\theta)$, $\hat{\alpha}_0(\theta)$ and $\hat{\alpha}_1(\theta)$ are $\sqrt{N}$ consistent and asymptotically jointly normally distributed. All the details about the procedures, the asymptotic distribution and the proofs can be found in the Appendix B.

3.2. Integration of nonparametric first step estimates

When the set of independent variables is discrete, the minimum distance framework provides an alternative estimation procedure. Buchinsky (1991) chapter 1 section 9 and Chamberlain (1994) derive and apply such an estimator for the exogenous case. In the presence of endogenous regressors, the idea consists in estimating the IV quantile regression separately in each cell and then to use the minimum distance framework to obtain $\sqrt{N}$ consistent and asymptotically normally distributed estimates of the coefficients. On the contrary of the method proposed in section 3.2, this approach can be directly extended to the case of continuous endogenous variables. Moreover, it is generally consistent with heteroscedasticity. On the other hand, the derivation of the asymptotic distribution of the estimator in the presence of continuous X is more tedious. Moreover, since the first step is estimated nonparametrically, only regions which satisfies the common support property ($0 < \Pr\left(D = 1 \middle| X\right) < 1$) can be used in this procedure.

Suppose that X_i comes from a discrete distribution so that there is a finite number, say *J*, of different possible vectors $X_{(j)}, j = 1,...J$. Chernozhukov and Hansen's estimator can be applied separately in each cell. Of course, only a constant $\alpha_j(\theta)$ and the quantile treatment effect (the coefficient on *D*) $\delta_j(\theta)$ are estimated. The asymptotic distribution of $\hat{\alpha}_j(\theta)$ and $\hat{\delta}_j(\theta)$ is directly derived from (A.2):

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}_j(\theta) - \alpha_j(\theta) \\ \hat{\delta}_j(\theta) - \beta_j(\theta) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{\Lambda_j(\theta)}{\Pr\left(X = x_{(j)}\right)}\right)$$

where $\Lambda_j(\theta)$ is equal to $\Lambda(\theta)$ defined in (A.2) with the exceptions that we condition all expected value on $X = x_{(j)}$ and the vector of regressors consists only of a constant term. We also note that $\left(\hat{\alpha}_j(\theta), \hat{\delta}_j(\theta)\right)$ is independent of $\left(\hat{\alpha}_{j'}(\theta), \hat{\delta}_{j'}(\theta)\right)$ for $j \neq j'$.

Recall that we assume that the conditional quantiles of *Y* given *X* are linear in each sector. That is,

$$\alpha_j(\theta) = X_{(j)}'\beta_0(\theta)$$

$$\alpha_j(\theta) + \delta_j(\theta) = X_{(j)}'\beta_1(\theta)$$

where the parameter vectors $\beta_0(\theta)$ and $\beta_1(\theta)$ are the same for $j = 1,...J$. Define *G* to be a $J \times k$ (with $J \geq k$) matrix with rows $X_{(1)},...,X_{(J)}$, $\alpha(\theta) = \left(\alpha_1(\theta),...,\alpha_J(\theta)\right)$, $\delta(\theta) = \left(\delta_1(\theta),...,\delta_J(\theta)\right)$ and $\hat{W}_s(\theta)$ is a $J \times J$ matrix that converges with probability one to $W_s(\theta)$, a positive-definite matrix, for $s = 0,1$. The minimum distance estimators of $\beta_0(\theta)$ and $\beta_1(\theta)$ are then defined by

$$\hat{\beta}_0(\theta) = \min_{\beta}\left(\hat{\alpha}(\theta) - G\beta\right)'\hat{W}_0(\theta)\left(\hat{\alpha}(\theta) - G\beta\right)$$

and $$\hat{\beta}_1(\theta) = \min_{\beta}\left(\hat{\alpha}(\theta) + \hat{\delta}(\theta) - G\beta\right)'\hat{W}_1(\theta)\left(\hat{\alpha}(\theta) + \hat{\delta}(\theta) - G\beta\right).$$

Then

$$\sqrt{N}\left(\hat{\beta}_s(\theta) - \beta_s(\theta)\right) \xrightarrow{d} N\left(0, \left(G'W_s(\theta)G\right)^{-1} G'W_s(\theta)\Omega_s(\theta)W_s(\theta)G\left(G'W_s(\theta)G\right)^{-1}\right),$$

for $s = 0,1$, where $\Omega_s(\theta)$ is a *J* diagonal matrix with the *j*th diagonal element equal to the variance of $\hat{\alpha}_j(\theta)$ and $\Omega_1(\theta)$ is a *J* diagonal matrix with the *j*th diagonal element equal to the variance $\hat{\alpha}_j(\theta) + \hat{\delta}_j(\theta)$. An efficient minimum distance estimator is obtained by setting $W_s(\theta)$ equal to a consistent estimator of Ω_s^{-1}.Note that if we estimate the same model as Chernozhukov and Hansen (without interaction term) the efficiently weighted minimum distance estimator is asymptotically identical to the estimator of Chernozhukov and Hansen with optimal instruments and weights.

If some of the exogenous variables are continuous, the problem becomes more complicated. We can first estimate α and δ nonparametrically at each observation using a locally weighted version of the instrumental quantile regression estimator and then use the minimum distance framework to obtain an estimate of the coefficients. Under some conditions on the kernel and the bandwidth, the second step is $\sqrt{N}$ consistent since we integrate over all observations and there are only a finite number of parameters to estimate. Chen, Linton and Van Keilegom (2003) provide the basic framework for deriving its asymptotic distribution. We will not follow this way in this paper. First, the Monte Carlo simulation in the following section show that the small sample properties of the minimum distance estimator are not really appealing. Second, the sample size we have in the application is probably too small for such an estimator and third, we would lose the good computational properties of the Chernozhukov and Hansen's estimator.

3.3. Finite sample properties of these estimators

In order to compare the performance of the two new estimators and to evaluate the costs in term of variance of allowing for interaction terms, we present the results of a Monte- Carlo simulation. The data generating process is the following:

$$X \sim b(0.5);\ Z \sim N(0,1);\ D \sim 1(0.5 - X + Z + \varepsilon < 0)$$

$$Y = \alpha + X\beta_x + D\beta_w + XD\beta_{xw} + (1+0.2D)U \qquad (6)$$

$$\varepsilon \sim t_3;\ U \sim t_3;\ Cov(\varepsilon, U) = 0.8.$$

In order to apply the minimum distance estimator in its simplest form, we only have a binary exogenous regressor. We consider two different sets of parameters:

model without interaction term: $\alpha = 0$, $\beta_x = 1$, $\beta_w = 1$ and $\beta_{xw} = 0$, model with interaction term: $\alpha = 0$, $\beta_x = 2$, $\beta_w = 1$ and $\beta_{xw} = -1$.

We set the number of observations to 100 and 400 and draw 10000 replications. In each replication, we apply the traditional quantile regression estimator (QR), the IQR of Chernozhukov and Hansen, the estimator described in section 3.1 which combines the sample selection correction of Buchinsky and the IQR of Chernozhukov and Hansen (SIQR) and finally the minimum distance estimator described in section 3.2 (MDIQR). IQR are estimated by searching on a grid between -1 and 3 with grid step of length 0.01. The selection equation is estimated by the smoothed maximum score estimator with a normal kernel and a bandwidth of 0.15 (see details in section 6.2). Since only 2 parameters must be estimated, the estimation is done by grid search. Since the results are similar for different quantiles, we present only the outputs for the median regressions. Tables 1 and 2 give the bias, the standard deviation (SD) and the mean squared error (MSE) for each estimator in the case without and with interaction term respectively.

From Table 1 we first note that QR is heavily biased for all coefficients, as a consequence of the high level of endogeneity presents in the data generating process. We also remark that the

Table 1: Monte-Carlo simulations, without interaction terms

Estimator	α (true value: 0)			β_x (true value: 1)			β_d (true value: 1)			β_{xd} (true value: 0)		
	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
	100 observations											
QR	-0.81	0.30	0.74	0.33	0.30	0.20	1.19	0.30	1.51			
IQR	0.03	0.46	0.21	-0.04	0.39	0.15	-0.02	0.55	0.30			
SIQR	0.02	0.54	0.30	-0.06	0.62	0.39	0.01	0.66	0.43	0.01	0.68	0.46
MDIQR	0.04	0.63	0.40	-0.05	0.73	0.53	-0.02	0.79	0.62	0.02	1.08	1.16
	400 observations											
QR	-0.81	0.14	0.67	0.33	0.15	0.13	1.18	0.15	1.42			
IQR	0.01	0.21	0.04	-0.01	0.17	0.03	0.00	0.25	0.06			
SIQR	0.01	0.25	0.06	-0.02	0.30	0.09	0.00	0.30	0.09	0.00	0.30	0.09
MDIQR	0.02	0.30	0.09	-0.02	0.34	0.12	-0.01	0.37	0.14	0.01	0.51	0.26

Note: 10000 replications. Estimators : QR : traditional quantile regression ; IQR : instrumental variable quantile regression estimator of Chernozhukov and Hansen (Appendix A) ; SIQR : instrumental variable quantile regression estimator using the sample selection correction procedure of Buchinsky (section 3.1) ; MDIQR : minimum distance instrumental variable quantile regression (section 3.2). The data generating process is given by equation (6).

standard deviation of IQR is about 50% higher than that of QR. Thus, the cost of allowing for endogeneity is not negligible but is necessary to obtain consistent results. In this first model, both estimators allowing for an interaction term have a higher standard deviation of about 20% compared to the IQR. The difference is particularly high for the estimation of β_x since only half of the observations are used to estimate β_x and the other half is used to estimate β_{xd}. This simulation also confirms the $\sqrt{N}$ convergence rate of all estimators since the standard deviation with 400 observations is about the half of the standard deviation with 100 observations.

Table 2 gives the results for the model with an interaction term. QR and IQR are biased since they disregard the endogeneity and the interaction term, respectively. The higher standard error of the SIQR and MDIQR is counterbalanced by the much lower bias which results in lower MSE, particularly with 400 observations. Now, if we compare both estimators allowing for interaction terms and endogeneity, we remark that SIQR is always more precise than MDIQR and that the difference is not negligible (between 20 and 60%). These differences are even higher if we take account of the fact that continuous regressors would complicate the task of the MDIQR but not of the SIQR.

Table 2: Monte-Carlo simulations, with interaction terms

Estimator	α (true value: 0)			β_x (true value: 2)			β_d (true value: 1)			β_{xd} (true value: -1)		
	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
	100 observations											
QR	-0.80	0.38	0.78	0.33	0.46	0.32	1.19	0.44	1.60	-0.01	0.59	0.35
IQR	0.42	0.52	0.44	-0.61	0.40	0.53	-0.54	0.57	0.61			
SIQR	0.02	0.54	0.30	-0.06	0.62	0.39	0.01	0.66	0.43	0.01	0.68	0.46
MDIQR	0.04	0.63	0.40	-0.07	0.72	0.52	-0.02	0.79	0.62	0.08	1.03	1.07
	400 observations											
QR	-0.82	0.19	0.71	0.36	0.23	0.18	1.20	0.22	1.50	-0.05	0.30	0.09
IQR	0.39	0.25	0.22	-0.58	0.19	0.37	-0.53	0.27	0.36			
SIQR	0.01	0.25	0.06	-0.02	0.30	0.09	0.00	0.30	0.09	0.00	0.30	0.09
MDIQR	0.02	0.30	0.09	-0.02	0.34	0.12	-0.01	0.37	0.14	0.01	0.51	0.26

Note: 10000 replications. Estimators : QR : traditional quantile regression ; IQR : instrumental variable quantile regression estimator of Chernozhukov and Hansen (Appendix A) ; SIQR : instrumental variable quantile regression estimator using the sample selection correction procedure of Buchinsky (section 3.1 and Appendix B) ; MDIQR : minimum distance instrumental variable quantile regression (section 3.2). The data generating process is given by equation (6).

4. Decomposition of differences in distribution

The most basic approach to explore the wage differential between groups or sectors involves estimating an earnings regression using pooled data for public and private sector employees and including a dummy variable for a worker's sector of employment. This specification can be estimated for the conditional mean or the conditional quantiles of the dependent variable. If the sector of employment is considered to be endogenous, the conditional mean of the log wage can be estimated by traditional instrumental variable. For quantile regression, recent developments presented in section 2 allow to correct for endogeneity. This simple dummy variable approach is pretty easy to estimate and the results are trivial to interpret since the "discrimination" part of the difference is the same for all observations at the same point of the distribution. However, a very strong restriction is implied by this specification: the returns to human capital characteristics are constrained to be equal across sectors. The effect of a worker's sector of employment is limited to be an intercept effect.

Since this restriction is often violated by the data, alternative methodologies have been proposed. The first step consists naturally in estimating the wage equation separately for each sector. Now, the discrimination is different at different point of the distribution of the covariates. A first possibility to present the results is to consider the expected wage rates or the quantiles of the wage distributions in the public and private sectors for reference individuals. Another common procedure consists in aggregating the results. The Oaxaca (1973) / Blinder (1973) decomposition is the best known decomposition procedure for models for the mean. It allows very easily to decompose the total difference into a part explained by different characteristics and a part explained by coefficients, which is often interpreted as discrimination. Decomposing differences in distribution is a more complex problem because the quantile of a linear function is not equal to linear function of the quantile contrarily to the mean.

Melly (2005b) proposes an intuitive procedure to decompose differences at different quantiles of the unconditional distribution. In a first step, the conditional distribution is estimated by quantile regression. In the second step, the conditional distribution is integrated over the range of the covariates. Formally, let $\hat{\beta} = \left(\hat{\beta}(\tau_1),...,\hat{\beta}(\tau_j),...,\hat{\beta}(\tau_J)\right)$ be the quantile regression coefficients estimated at J different quantiles $0 < \tau_j < 1$, $j = 1,...,J$. Integrating over all quantile regression and over all observations, a natural estimator of the θ^{th} unconditional quantile of the dependent variable is given by

$$Q(\theta, X, \beta) = \inf\left\{q : \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(\tau_j - \tau_{j-1}\right)1\left(x_i\hat{\beta}(\tau_j) \leq q\right) \geq \theta\right\}.$$

Melly (2005b) shows that this estimator is consistent and asymptotically normally distributed. Consistent estimators of the variances are also proposed. Now, we can estimate counterfactual distributions by replacing the estimated coefficients or the distribution of characteristics in a sector with the estimated coefficients or the distribution of characteristics in the other sector. It is thus possible to separate the difference at each quantile of the unconditional distribution into a part explained by coefficients and a part explained by characteristics:

$$\begin{aligned} &Q\left(\theta, X^{pub}, \beta^{pub}\right) - Q\left(\theta, X^{priv}, \beta^{priv}\right) \\ &= \left[Q\left(\theta, X^{pub}, \beta^{pub}\right) - Q\left(\theta, X^{pub}, \beta^{priv}\right)\right] + \left[Q\left(\theta, X^{pub}, \beta^{priv}\right) - Q\left(\theta, X^{priv}, \beta^{priv}\right)\right] \end{aligned}$$

where the first bracket represents the effect of differences in coefficients (discrimination) and the second bracket represents the effect of differences in the distribution of characteristics (justified differential).

In presence of endogenous sector choice, the same procedure can be used with the coefficients estimated by the procedures proposed in section 3. Then, we can estimate the wage distributions that we would observe without the sample selection bias. Thus, the difference between the quantiles of the unconditional distribution in the public sector and the quantiles of the unconditional distribution in the private sector can be decomposed into three components: effects of sample selection, effects of differences in coefficients and effects of differences in the distribution of characteristics.

5. Data, descriptive statistics and instruments

The analysis in this paper draws on data from the German Socio-Economic Panel (GSOEP)[5] for the year 2003. It would be interesting to use the panel structure of the data to estimate a fixed effect model. Unfortunately there is not enough movement between the public sector and the private sector to obtain useful results. Therefore we concentrate in this paper on the last wave of the panel and we control for endogeneity of the sector choice by instrumental variable methods. After the reunification, the panel was extended to include the eastern part of Germany, but we focus here on West Germany because undeniable economic differences subsist between East and West Germany. Since many public sector jobs are not open to foreign nationals, the analysis is based on the subsample of Germans only. Furthermore, the sample is restricted to include only men who were between 17 and 65 years old and were in full-time or part-time employment. As the sample includes only wage earners, the results must be interpreted conditional on the selected sample. However since we concentrate on males, we can hope that this selection bias is not important. Finally, all observations with a missing value for one of the variables have been excluded. The final dataset has 3125 observations.

Table 3 defines the variables we use for our empirical analyses. *Y*, the dependent variable, is *Lnghearn*, the logged gross hourly wage. *X*, the vector of regressors assumed to be exogenous, contents a quadratic in potential experience and 5 educational dummies[6]. *D*, the endogenous variable, is *Psect*, a dummy variable equal to 1 if the person is employed in the public sector and 0 if she is employed in the private sector. We do not

[5] For an English language description of the GSOEP see SOEP Group (2001).

[6] The exogeneity of education is rejected by Dustmann and Van Soest (1998). However, to avoid computational difficulties and presumably very high variances of the estimates, we do not consider this possibility in this paper.

Table 3: Definition of the variables

Variable	*Description*
Ghearn	Gross hourly earnings from employment. Gross hourly wage are derived by dividing gross monthly earnings by monthly actual hours worked.
Lnghearn	The natural logarithm of Ghearn.
Expr	Number of years of potential work experience the individual has accumulated. It is measured by min(age-schooling-6, age –18).
Ed level	Ordered variable on education:
Ed level 1	Dummy; 1 if no degree or basic or intermediate schooling with no training.
Ed level 2	Dummy; 1 if basic schooling with apprenticeship.
Ed level 3	Dummy; 1 if intermediate schooling with apprenticeship.
Ed level 4	Dummy; 1 if high school (Abitur or Fachabitur) with no training or with apprenticeship.
Ed level 5	Dummy; 1 if high school with technical school or polytechnic.
Ed level 6	Dummy; 1 if university.
Psect	Dummy; 1 if employed in the public sector.
Fcivil	Dummy, 1 if father civil servant at the time the respondent was 16 years old.
Fblue	Dummy, 1 if father blue collar at the time the respondent was 16 years old.
Fself	Dummy, 1 if father self employed at the time the respondent was 16 years old.
Fwhite	Dummy, 1 if father white collar at the time the respondent was 16 years old.
Mnwork	Dummy, 1 if mother did not work at the time the respondent was 16 years old..

distinguish between civil servants (Beamte) and other public sector employees since pay scales are the same and apply to all public sector workers at the federal, state and local level. Table 4 presents descriptive statistics for public and private sector employees. Means of the relevant variables show that average hourly earnings are higher in the public sector than in the private sector. They also show that public sector employees are, on average, better educated than private sector employees. For instance, 22.7% of the employees in the public sector have achieved a university degree (*Ed level 6*), while they are only 13% in the private sector. Public sector employees have acquired more labor market experience, too. These differences in work experience and education may explain the higher average wages of public sector employees.

A first visual summary of the public and private sector wage distributions is provided in Figure 1. The density functions were estimated using an Epanechnikov kernel estimator and

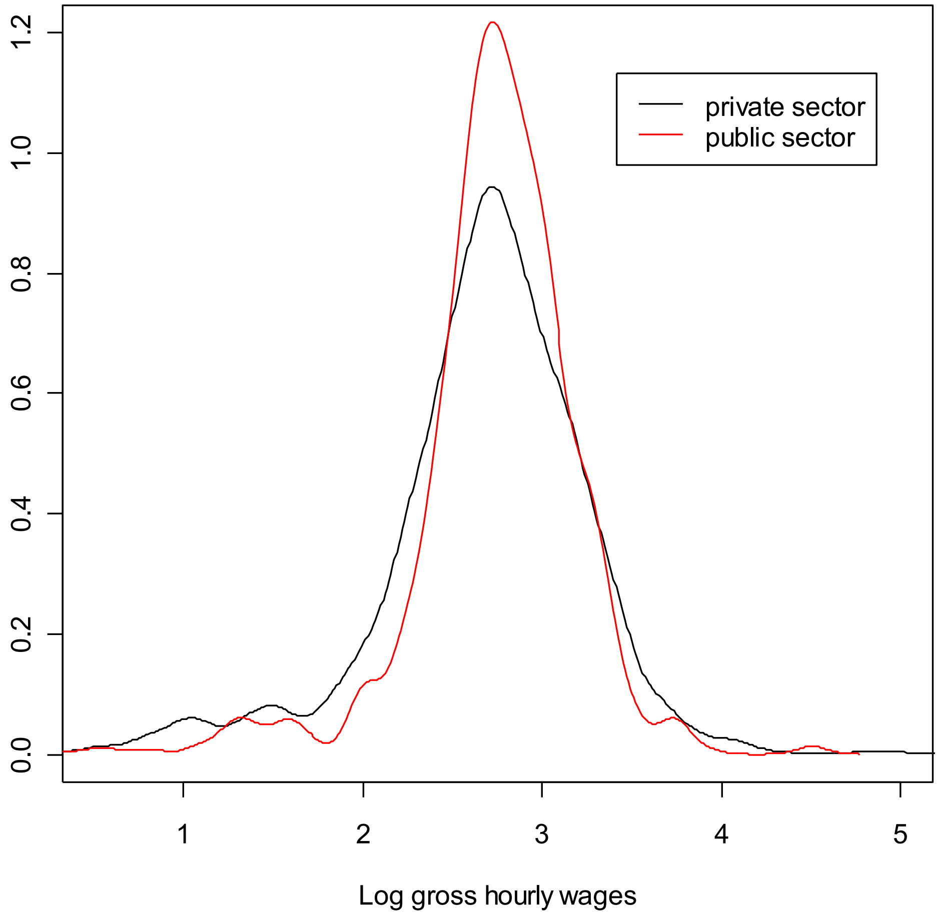
Table 4: Descriptive statistics, means

Variable	All	Public Sector	Private Sector
Lnghearn	2.693	2.745	2.677
Expr	22.09	24.15	21.47
Education:			
Ed level 1	9.7%	5.7%	10.9%
Ed level 2	30.1%	22.3%	32.4%
Ed level 3	24.7%	26.4%	24.2%
Ed level 4	8.4%	9.3%	8.1%
Ed level 5	11.9%	13.5%	11.4%
Ed level 6	15.3%	22.7%	13%
Fcivil	10.3%	16.2%	8.6%
Fblue	39.9%	35%	41.4%
Fself	12%	12.6%	11.8%
Fwhite	21.6%	23.2%	21.2%
Mnwork	21.2%	25.2%	20%
Number of observations	3125	717	2408

the bandwidth was chosen according Silverman's rule of thumb (1986). It can be seen from this figure that the distributions are quite distinct between sectors. The public sector earnings distribution is characterized by a higher density function around the mode and a lower dispersion. The public sector earnings distribution lies "within" the private distribution. Public sector employees at the 10^{th} quantile of the public sector earnings distribution enjoy an earnings advantage over private sector employees at the same point in the private sector distribution of wages; but the reverse holds for employees at the 90^{th} quantile of the public sector and private sector earnings distribution. With "higher floors" and "lower ceilings", the public sector compresses the unconditional wage distribution.

Given that there is a choice being made by workers whether to work in the public or private sector, there is the potential for sample selection bias. To correct for endogenous sector choice, nonparametric identification requires exclusion restrictions. In many studies, the data is not rich enough to provide appropriate instruments and identification assumptions are

Figure 1: Kernel density estimates of the wage distributions



Note: Density functions estimated using an Epanechnikov kernel estimator and bandwidths chosen according Silverman's rule of thumb.

sometimes doubtful. For example, different measures of education have been used in the wage equation and in the selection equation or age is used in one equation and experience in the other. The GSOEP is a rich dataset that contains a large range of background variables usually not available in other studies. We will use 5 variables defined in Table 3 related on parents' occupational status. Dustmann and Van Soest (1998) have used very similar exclusion restrictions. The most important instrument is *Fcivil*, a dummy variable that is equal to 1 if the father was a civil servant at the time the employee was 16 years old. Table 4 shows high correlations between the instruments and the public sector status. For instance, if the father worked in the public sector, his son will also work in the public sector with a probability of 36%. If the father did not work in the public sector, the probability is only of 21%.

Our motivation for using these instruments is that children learn through imitation of adults living in their neighborhood. The image the parents shown to the child is something that he will use as a base for his own growth. For a son, the father sets an example, a reference to imitate. The exclusion restriction (1) could be violated, for instance, if the father had better relationships in his sector of employment and his son could benefit from his relationships to increase his wage. As an indication against this hypothesis, we find that the sector choice of the father is not important for his daughter but is strongly significant for his son. Similarly the sector choice of his mother plays no role in the occupational choice of her son. If the reason for the correlation between the sector choices of children and parents was the relationships the father has in his sector, we would find a positive correlation also for girls. Finally, note that we have 5 instruments for a single endogenous variable; we will use the overidentification to test the exclusion restriction in section 6.3.

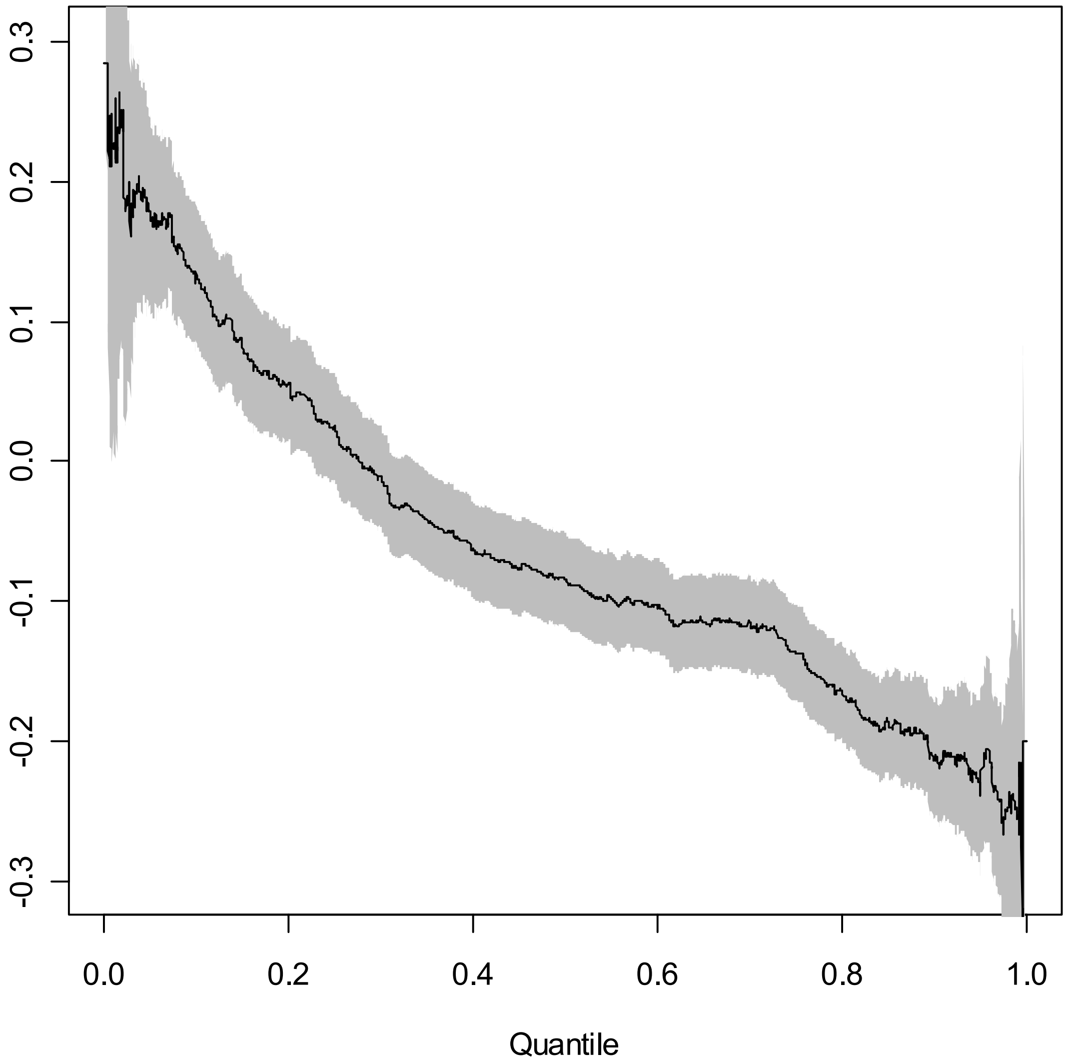
6. Empirical results

6.1. Exogenous sector choice

As a benchmark, we first estimate the public-private sector wage differential using traditional quantile regression methods, assuming that the sector choice is exogenous. The first method used is the simple dummy variable approach. We have regressed the logged wage on X and on the public sector dummy with traditional quantile regression. The estimated public sector gap as a function of θ is plotted in Figure 2 with a 95% confidence interval. All standard errors in this paper, if the contrary is not explicitly stated, were estimated by the sample analogs of the asymptotic results[7]. The densities were estimated by the method of Powell (1984) using a normal kernel and a bandwidth following the Bofinger's (1975) rule. The estimated

[7] That means that we use the proposals of Powell (1984) for traditional quantile regression, of Buchinsky (1998) for the sample selection correction, of Chernozhukov and Hansen (2005b) for the IQR and of Melly (2005b) for the decomposition procedure.

Figure 2: Public sector wage "premium" at different quantiles



Note: coefficient on the public sector dummy variable estimated by traditional quantile regression with a 95% confidence interval.

coefficients for the median regression are given in Table 5. At this point of the distribution public sector employees earn 8.4% less than private sector employees with the same characteristics and this coefficient is significantly different from zero. The results of Figure 2 show that the public sector compresses the wages by giving a positive premium at the low end of the conditional distribution and a significant negative premium at the upper tail of the distribution.

These results are correct only if the returns to human characteristics are the same in both sectors. To test this restriction we have estimated a fully interacted model where all characteristics are interacted with the public sector dummy and we have tested if the interaction terms are significantly different from zero. The null hypothesis is clearly rejected for most quantiles and is definitely rejected for the whole quantile regression process. Therefore, we have estimated 100 quantile regressions separately in each sector. The results of the median regressions, given in the 4th and 5th columns of Table 5, show that returns to education are generally higher in the private sector, the linear term of the polynomial in experience is higher in the private sector but the contrary happens for the quadratic term.

Then, using the procedure described in section 4, we have decomposed the differences between the quantiles of the unconditional distributions into a part explained by different distributions of characteristics and a part explained by different coefficients (could be

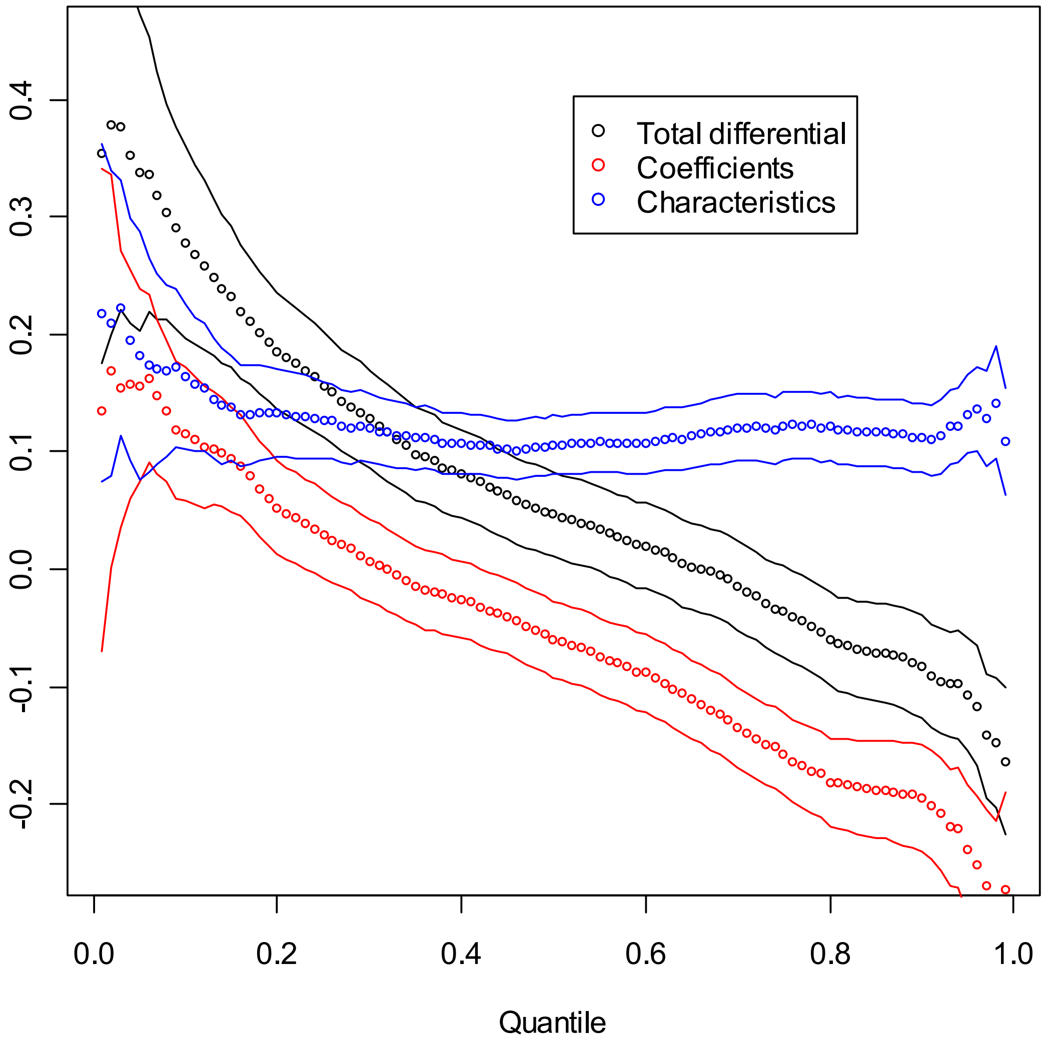
Table 5: Median regression using different estimators

	RQ	IRQ1	IRQ2	RQ public	RQ private	SIRQ public	SIRQ private
Constant	1.614**	1.522**	1.663**	1.715**	1.572**	2.008**	1.582**
	(0.056)	(0.064)	(0.061)	(0.132)	(0.062)	(0.245)	(0.066)
Expr	0.062**	0.063**	0.057**	0.051**	0.064**	0.052**	0.06**
	(0.004)	(0.005)	(0.005)	(0.008)	(0.005)	(0.008)	((0.005)
Expr^2	-1e-3**	-0.001**	-9e-4**	-7e-4**	-1e-3**	-7e-4**	-1e-3**
	(9e-5)	(1e-4)	(9e-5)	(2e-4)	(1e-4)	(2e-4)	(1e-4)
Ed level 2	0.198**	0.26**	0.181**	0.071	0.236**	0.1	0.276**
	(0.041)	(0.047)	(0.043)	(0.092)	(0.047)	(0.1)	(0.061)
Ed level 3	0.317**	0.32**	0.251**	0.21*	0.359**	0.189*	0.298**
	(0.042)	(0.049)	(0.0477)	(0.091)	(0.048)	(0.094)*	(0.056)
Ed level 4	0.34**	0.322**	0.251**	0.287**	0.346**	0.261*	0.23**
	(0.053)	(0.07)	(0.069)	(0.104)	(0.065)	(0.113)	(0.079)
Ed level 5	0.587**	0.587**	0.523**	0.438**	0.655**	0.393**	0.56**
	(0.045)	(0.055)	(0.053)	(0.096)	(0.054)	(0.114)	(0.079)
Ed level 6	0.709**	0.659**	0.606**	0.527**	0.775**	0.494**	0.698**
	(0.045)	(0.069)	(0.066)	(0.094)	(0.054)	(0.101)	(0.068)
Psect	-0.084**	0.31	0.224**				
	(0.02)	(0.237)	(0.217)				

Note: column 1: quantile regression, column 2: instrumental variable quantile regression with 5 instruments, column 3: instrumental variable quantile regression with efficient instrument and weights estimated by OLS projections, column 4: quantile regression in the public sector, column 5: quantile regression in the private sector, column 6: SIVQR in the public sector, column 7: SIVQR in the private sector. *: significant at the 5% level, **: significant at the 1% level. Standard errors are given in parenthesis.

interpreted as premium or discrimination). Figure 3 plots the decomposition results with a 95% confidence interval for all estimates. The compression of the unconditional public sector wages distribution can be seen by looking at the total differential. The 10% quantile of the public sector wage distribution is higher than the 10% private sector wages distribution but the contrary holds at the 90% quantile. This is only another way of presenting the results of Figure 1. The part explained by characteristics is significantly positive, reflecting the fact that the public sector employees are better educated and have more experience than private sector employees. We cannot reject the hypothesis that the part explained by characteristics is constant across the distribution. Therefore, the higher wage dispersion in the public sector is not caused by higher dispersion of the characteristics. Finally, the part explained by

Figure 3: Decomposition of public private sector wage differential at different quantiles



Note: 95% confidence intervals are delimited by the lines.

coefficients is very similar to the results of the dummy variable approach in Figure 2. The premium is significantly negative at the median and decreases monotonically from the low end to the high end of the distribution. Similar results have been found by Melly (2005a). However, we should recall that these results were obtained assuming exogenous sector choice. This unlikely to be satisfied assumption will be suppressed in sections 6.3 and 6.4.

6.2. Choice between private and public sector

To describe the selection process between both sectors and as a first step estimation for the sample selection correction procedure of Buchinsky (1998), we estimate the probability of working in the public sector conditionally on X and Z by a logit. Since the logit depends for consistency heavily on the distributional assumption, we also estimate the sector choice equation by smoothed binary quantile regression (Kordas 2005). Smoothed binary quantile regression was preferred to Klein and Spady's (1993) or Ichimura's (1993) estimators because it allows for arbitrary heteroscedasticity. Moreover, it imposes the same type of restrictions as quantile regression in the other steps of the estimation (conditional quantile restrictions).

Maximum score estimation, developed and studied by Manski (1975 and 1985) is equivalent to median regression applied to the binary choice model. Kim and Pollard (1990) established a $O\left(n^{-1/3}\right)$ rate of convergence for the estimator and showed that its limited distribution cannot be used for inference because it depends on unknown nuisance parameters. To overcome these shortcomings, Horowitz (1992) proposed smoothing the objective function. Kordas (2005) extends results regarding smoothed median regression to general smoothed binary quantile regression. While quantile regression is interesting because it provides a more complete picture of how covariates affect the dependent variable, binary quantile regression is even more interesting because it allows to identify the model in cases where the median parameters are not identified. If both are identified, binary quantile regression can be much more efficient than maximum score estimator. Recall that the maximum score estimator uses effectively only the observations for which $\Pr\left(D=1|X,Z\right)=0.5$. In the public-private sector application, 22.9% of the employees in the sample work in the public sector. Using the logit results, only 129 observations or 4% of the sample have a probability of working in the public sector higher than 50%. Therefore, the (smoothed) maximum score estimator will have a very high variance since it depends on very few observations. A binary quantile regression for a quantile higher than the median will have a lower variance. We choose $\theta = 1-\overline{D} = 0.771$.[8] Since the model is identified only up to scale we normalize the coefficient of experience to 1. The smoothed binary quantile regression estimator is defined as the solution to the problem

$$\hat{\alpha}(\theta) = \arg\max_a \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - (1-\theta)\right)K_h\left(\frac{X_i{}'a}{\sigma_n}\right)$$

where $K_h(\cdot)$ is the integral of a h^{th} order kernel function and σ_n is a bandwidth converging to 0 as $n \to \infty$. Given the conditions of Horowitz (1992) or Kordas (2005), $\hat{\alpha}(\theta)$ is asymptotically normally distributed. The fastest possible rate of convergence of $\alpha(\theta)$ is $O\left(n^{-h/(2h+1)}\right)$. We use the fourth order kernel proposed by Horowitz (2002). Using the plug-in method of Horowitz (1992) to estimate the optimal bandwidth, we obtain an optimal bandwidth near 1 for a large range of starting values. To remove the asymptotic bias, we undersmooth and choose a bandwidth of 0.1. The variance of the estimates was obtained by bootstrapping the results 200 times[9]. The objective function to be maximized has many local maxima and requires a global optimization algorithm. We use the algorithm developed by Sekhon and Mebane (1998) which combines evolutionary algorithm methods with a derivative-based method. The genetic algorithm yields a value of $\hat{\alpha}(\theta)$ that is sufficiently

[8] If the error term distribution is unimodal and symmetric and if we abstract from the presence of regressors, the optimal quantile will be $1-E[D]$. In realistic situations, it is more complicated to find the optimal quantile but $1-E[D]$ can be considered as a reasonable approximation.

[9] Horowitz (2002) proves the validity of the bootstrap.

near the global maximum and then a quasi-Newton method finds the local maximum since the objective function is smooth.

The results of the logit and smoothed binary quantile regression estimations are given in Table 6. The coefficients of the logit and of the smoothed binary quantile estimator are not fundamentally different with the exception of the constants because of the different normalizations. The standard errors of the logit estimates are generally slightly lower than those of smoothed quantile estimates, as expected, but the differences are not huge. The probability to work in the public sector increase with (potential) experience and education. One of the fundamental assumption of the instrumental variable quantile regression estimator is the presence of at least one instrument having an effect on the endogenous variable. We can test this assumption using the results of the smoothed binary quantile regression and we find three significant instruments: *Fcivil* is significantly different from zero at the 1 per mil, *Mnwork* and *Fwhithe* are significant at the 5% level. The Wald test for testing the hypothesis that the coefficients of all instruments are equal to zero gives a value of 29.157 rejecting the null hypothesis at all sensible significance levels.

Table 6: Estimation of the selection equation, dependent variable: *psect*

	Logit		Smoothed binary quantile	
	Coefficient	Std. error	Coefficient	Std. error
Constant	-67.713***	6.431	-47.007***	5.867
Expr	1**	0.449	1***	
Expr^2	-0.006	0.009	0.002	0.008
Ed level 2	-0.91	4.945	-16.613**	7.719
Ed level 3	12.562**	4.905	14.098***	4.868
Ed level 4	18.54***	5.665	21.628***	5.808
Ed level 5	14.459***	5.382	33.975***	8.046
Ed level 6	24.435***	5.13	19.507***	5.139
Fcivil	17.427***	3.993	30.834***	7.161
Mnwork	5.574**	2.621	7.045**	3.546
Fblue	4.496	3.36	6.913	4.205
Fself	5.445	4.161	4.142	5.249
Fwhite	6.54*	3.574	8.498**	4.174

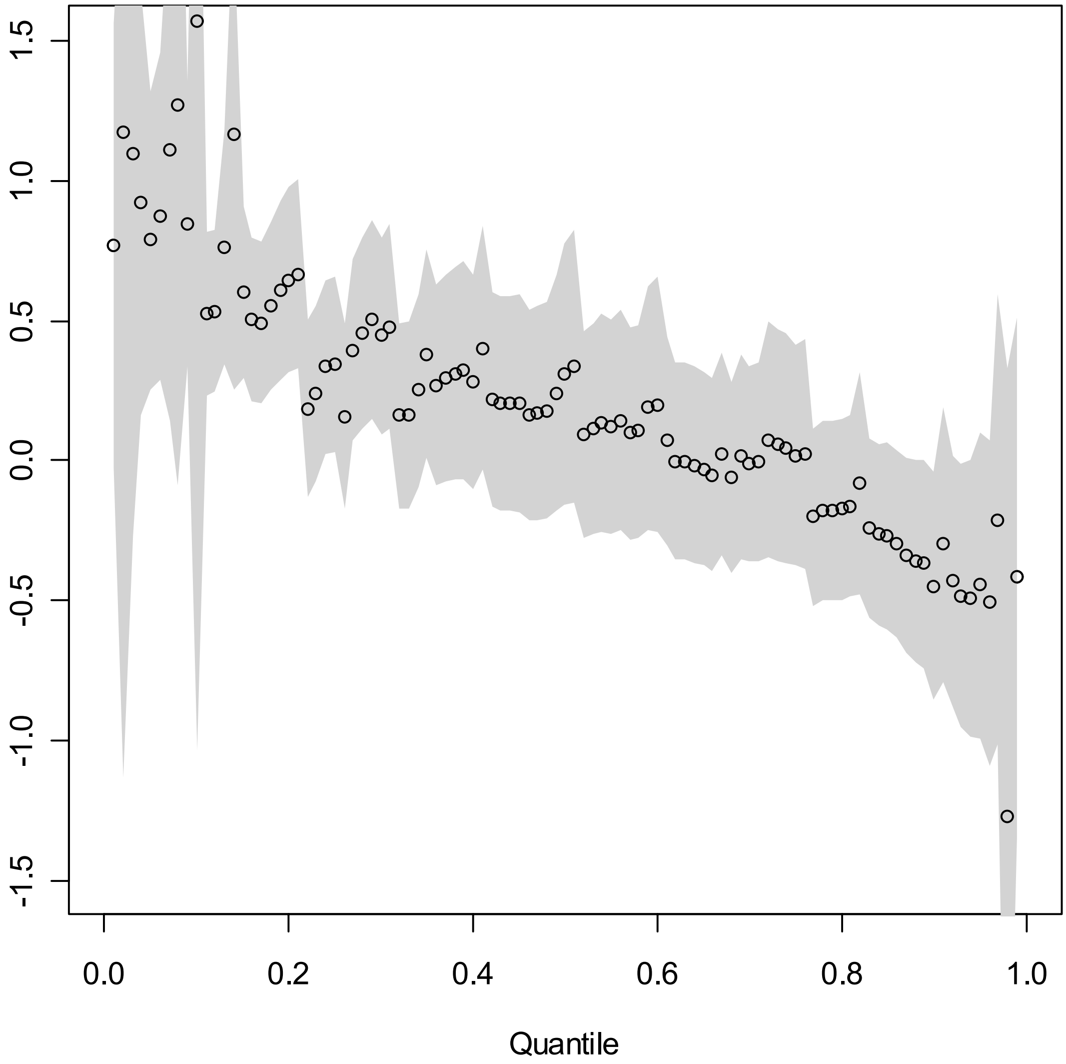
Note: Standard errors for the smoothed binary quantile estimator are obtained by bootstrapping the results 200 times. *: significant at the 10%, **significant at the 5%, ***: significant at the 1%

6.3. Endogenous dummy variable

In this section we correct for the endogeneity of the sector choice by using the estimator of Chernozhukov and Hansen (2004b and 2005b) presented in Appendix A. The crucial assumptions of their estimator (apart from the parametric restriction) are the presence of instruments and the exclusion of these instruments from the outcome equation. In section 6.2 we find that at least three instruments have an effect on *D*, confirming the first assumption. Since we have 5 instruments for a single endogenous variable, we can also partially test the second assumption. If we choose the weighting matrix $\mathrm{A}(\theta,\alpha)$ to be the inverse of the asymptotic covariance matrix of $\sqrt{n}\left(\hat{\gamma}(\theta,\alpha)-\gamma(\theta,\alpha)\right)$, the objective function of the IQR is asymptotically $\chi^2_{\dim(\gamma)}$-distributed under the null-hypothesis that the exclusion restrictions are satisfied. We estimate the instrumental quantile regression objective function at the 99 percentiles. For none of the percentile we can reject the null-hypothesis at the 1% significance level. For instance, the value of the objective function is 6.74 at the median.

We first present the simplest version of the estimator and then we try to estimate the efficient instruments and weights. First, we do not weight the observations and we chose $\mathrm{A}(\theta,\alpha)$ to be the inverse of the asymptotic covariance matrix of $\sqrt{n}\left(\hat{\gamma}(\theta,\alpha)-\hat{\gamma}(\theta,\alpha)\right)$. The estimation procedure consists simply in running a series of standard quantile regressions of $Y-D\alpha$ on covariates *X* and instrument *Z* over a grid of α. The parameter space for α was taken to be between -2 and 2 for $\theta<0.2$ or $\theta>0.8$ and between -1 and 1 for the other quantiles. We used an equally spaced grid with step size of 0.001. Figure 4 plots the coefficients on the public sector dummy at each percentile. If we compare these results with the results of section 6.1, the premium is about 40% higher if we correct for endogeneity. While the differential was negative over the major part of the distribution with quantile regression, it is now positive over 75% of distribution. The correction for endogeneity of the sector choice inverts the

Figure 4: Public sector wage premium using instrumental quantile regression

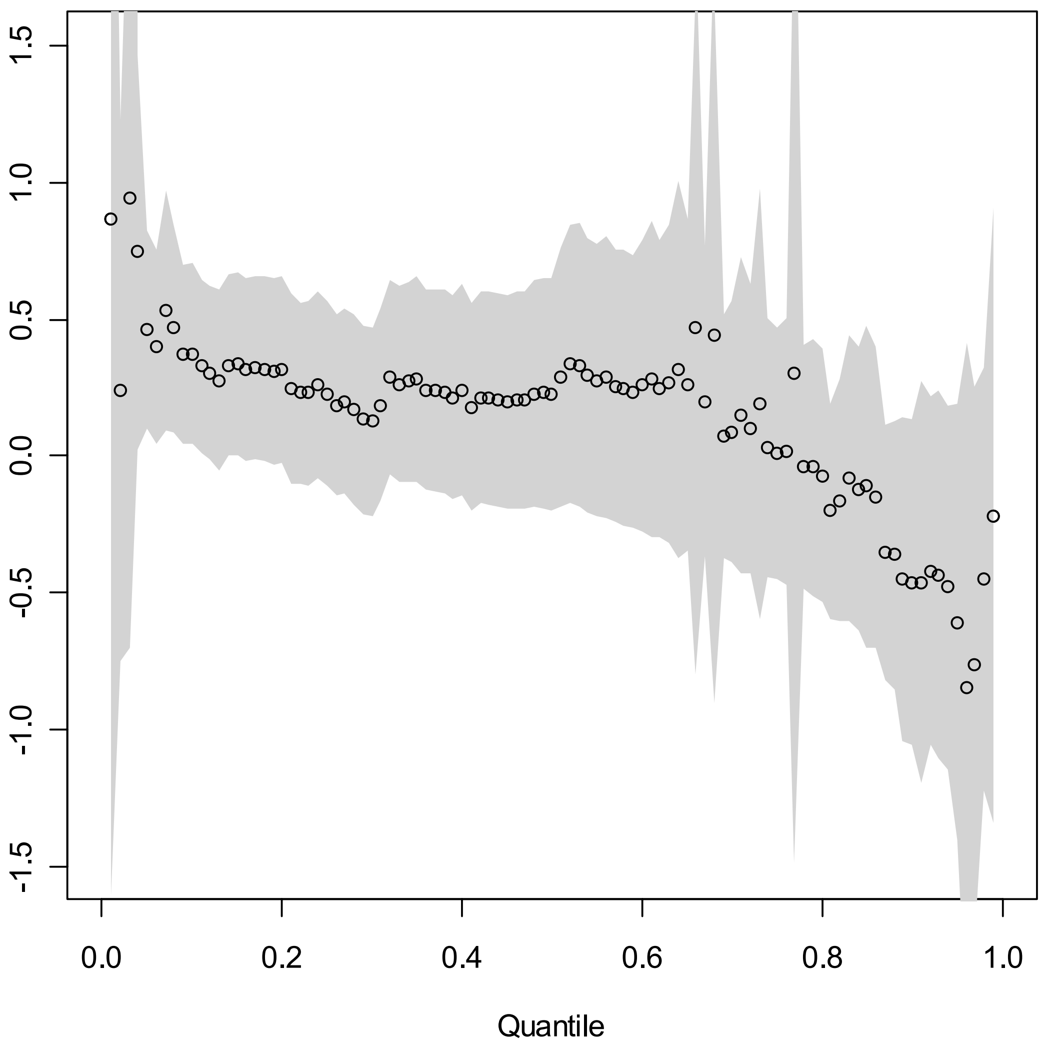


Note: Instrumental variable quantile regression with 5 instruments. Coefficient on the public sector dummy variable with a 95% confidence interval.

conclusion: the majority of public sector employees is over- and not underpaid. If we consider the evolution of the premium when we let θ vary between 0 and 1, we obtain a similar picture to Figure 2. The premium declines more or less monotonically from high positive values at the lower end of the distribution to negative values at the higher end of the distribution. The differences are even more pronounced with estimates ranging from 1.5 to –1.5. However, given the variance of the estimates, these extreme results should be kept with caution. In any case, the compression of the pay distribution by the public sector remains after correcting for endogeneity and is maybe even accentuated. Thus, the different distributions of wages in both sectors are not caused by different distributions of unobserved ability.

As explained in Appendix A, an efficient estimator can be obtained by using estimated instruments and weights. Given the small sample size, we estimate the instruments and weights parametrically by using least squares projections. Note however that inconsistent estimation of the instruments do not bias the results but only entails the efficiency of the estimator. The results of the public sector wage gap and the 95% confidence interval are plotted in Figure 5. They are not fundamentally different from those in Figure 4. The most notable difference is that the estimates vary less among the distribution and are less extreme for low and high quantiles, which is probably more credible. Slightly disappointing is the fact

Figure 5: Public sector wage premium using instrumental quantile regression



Note: Instrumental variable quantile regression with efficient instrument and weights estimated by OLS projections. Coefficient on the public sector dummy variable with a 95% confidence interval.

that the standard errors of the estimates do not really decrease by using the efficient estimator. For the coefficients on the public sector dummy, the standard errors are almost 10 times higher than for the exogenous quantile regression. Therefore, it is not possible to draw significant conclusions for a single quantile.

However, we can draw clearer conclusions by considering the whole instrumental quantile regression process. Chernozhukov and Hansen (2005b) propose inference procedures to evaluate the impact of the treatment on the entire distribution of outcomes. They suggest a subsampling procedure to compute asymptotically valid critical values for these tests. They propose a method of score resampling but we prefer to recompute the estimates in each replication, which avoid the estimation of conditional densities. We use the Smirnov-Cramer-Von-Misses statistic[10] and we estimate Anderson-Darling weights by subsampling. We estimate the critical values by constructing 1000 replications with 250 observations drawn without replacement[11] and estimate the instrumental quantile regression process on $\tau \in [0.1, 0.9]$.

Table 7: Tests on the instrumental quantile regression process

Null hypothesis	P-value with instruments weighted by the inverse of the variance	P-value with the optimal instruments and weights estimated by least squares projections
No effect: $\alpha(\cdot)=0$	<0.1%	0.1%
Constant effect: $\alpha(\cdot)=\alpha$	<0.1%	1.5%
Exogeneity: $\alpha(\cdot)=\alpha_{QR}(\cdot)$	<0.1%	<0.1%
Dominance: $\alpha(\cdot)\geq 0$	36.3%	45.1%
Dominance: $\alpha(\cdot)\leq 0$	<0.1%	<0.1%

Note: The statistics and critical values are computed using the method of Chernozhukov and Hansen (2002) using the Smnirnov-Cramer-Von Misses statistic. 1000 replications with 250 observations without replacement were constructed.

[10] The Kolmogorov-Smirnov statistic gives similar results

[11] They recommend choosing a block size of $kn^{2/5}$ with k between 3 and 10. $k=10$ gives a block size of 250.

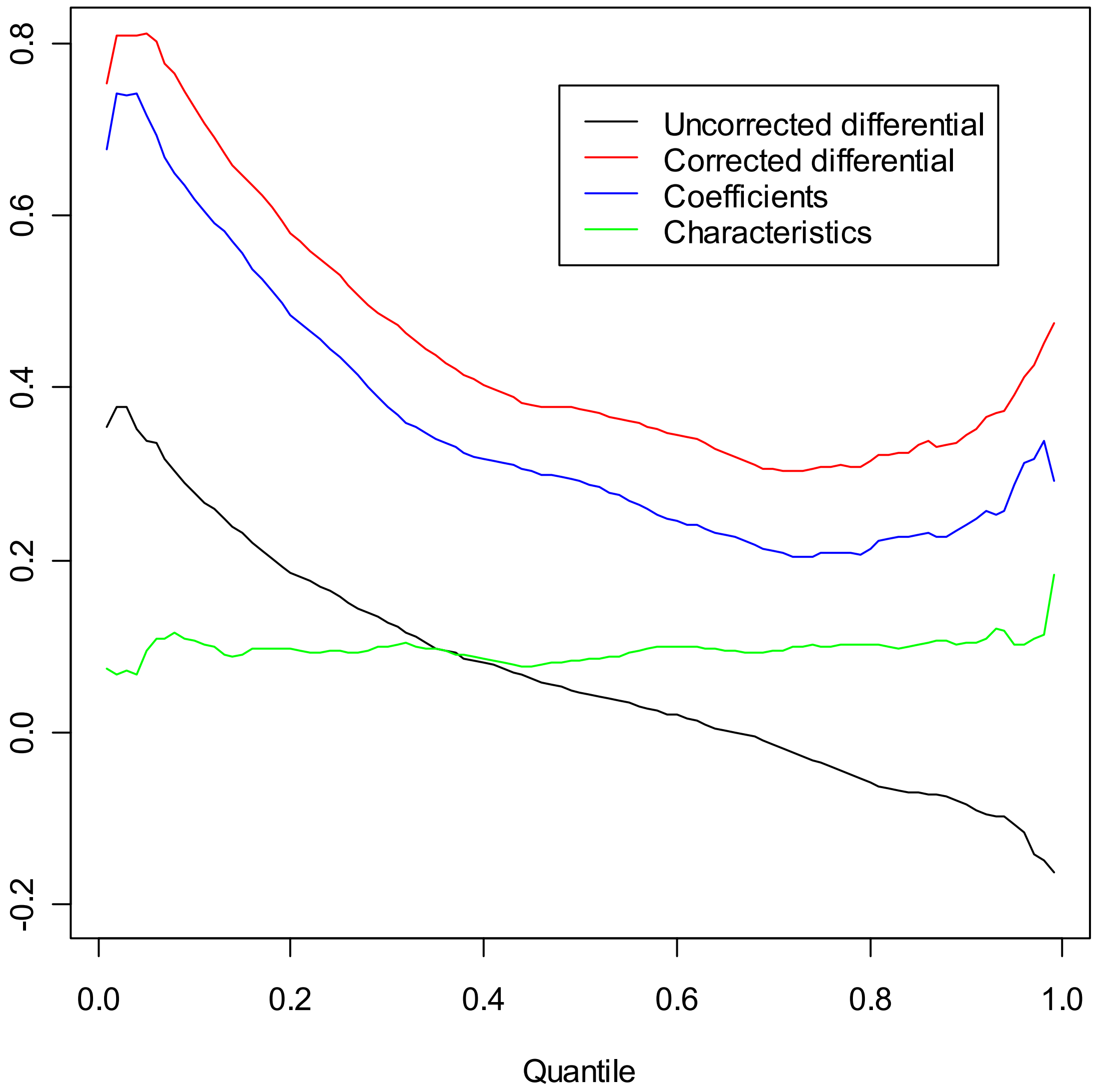
The p-values for 5 hypotheses are given in Table 7. As expected, we can reject the hypothesis that there is no difference between both sectors. The tests also strongly reject the null hypothesis of a constant effect, which was taken to be the weighted trimmed mean on $\tau \in [0.1, 0.9]$. Furthermore, the tests reject strongly the hypothesis of exogeneity, confirming the need to instruments for the sector choice. This confirms the results of Figures 4 and 5 and show that there is positive selection into the private sector. Finally, we reject the hypothesis that the wage distribution in the private sector dominates the distribution in the public sector but we cannot reject the opposite.

6.4. Endogenous sector choice with fully interacted covariates

The estimators used in section 6.3 assume that the returns to characteristics are the same in both sectors. In the exogenous case, it was shown that this restriction is not satisfied. Therefore, we apply now the estimators proposed in section 3. Given the results of the Monte-Carlo simulation, we concentrate on the SIQR of section 3.1. Results for the MDIQR (not presented) are principally similar but have higher variances.

The sample selection procedure of Buchinsky (1998) is used to estimate the slope coefficients. The index was estimate by binary quantile regression (section 6.2) and we approximate the bias term with a second-order power series expansion of the inverse Mill's ratio. Then the constants are estimated by the Chernozhukov and Hansen (2004b) estimator. For the possible values of α we use the same grids as in section 6.3. The instrument was taken to be the least squared projection of D on X and Z. The results of the median regressions, given in the 6th and 7th columns of Table 5, show that returns to education and experience are not the same in both sectors. Formal tests reject the null hypothesis of equal slopes at the 1% significance level. Similar results arise from the other quantiles, although the p-value is somewhat higher in the more extreme parts of the distribution. A joint test of equal slopes in both sectors at the 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles reject the null hypothesis at the 0.1% level[12]. Returns to education are generally lower in the public sector. Thus not only within-group inequality is lower in the public sector but also between-group inequality. Returns to (potential) experience are also higher in the private sector for younger employees. However, since the function is more concave in the private sector, the situation is inverted at the end of the work life (more than 27 years of experience).

Figure 6: Decomposition of public private sector wage differential correcting for endogeneity



Note: The uncorrected differential is the observed differential between the quantiles of the public sector wage distribution and the quantiles of the private sector wage distribution. The corrected differential is the differential that would prevailed if the sector choice was exogenous. The effect of characteristics is the effect of the different distributions of characteristics. The effect of coefficients can be interpreted as discrimination. All coefficients used to estimate the counterfactual distribution were estimated by the SIQR estimator.

[12] Since we do not have developed the results for the SIQR process until now, we cannot apply a procedure similar to Chernozhukov and Hansen (2005b). This is left for further research. The estimation of the variance by a bootstrap with 200 replications does not change the significance of the tests.

We have estimate the coefficients vectors corrected for endogeneity at 100 different quantiles uniformly distributed between 0 and 1 ($\theta = 0.005, 0.015, ..., 0.995$). The procedure described in section 4 allows the estimation of the potential wage distributions in both sectors and of the counterfactual distribution that would prevail if public sector employees were paid like private sector employees. Figure 6 plots the results. No confidence intervals are plotted to avoid to surcharge the figure. At the median, the standard errors of the estimates are 1.8%, 14.27%, 14.91% and 1.91% for the uncorrected differential, corrected differential, effects of coefficients and effects of characteristics, respectively. The uncorrected differential is taken from Figure 3 and represents the observed differences between the quantiles of the wage distribution in the public sector and in the private sector. The corrected differential is the differential that we would observe if the employees sorted randomly between sectors conditionally on their characteristics. We note that the corrected differential is much higher than the uncorrected one, showing that there is positive selection into the private sector and negative selection into the public sector. Now, we decompose the corrected differential into the part explained by different characteristics distributions and the part explained by different coefficients (often interpreted as discrimination). The effects of characteristics are positive and stable across the distribution, as they were assuming exogeneity. The effects of coefficients decreases as we move on the wage distribution but remain positive at all quantiles, indicating that a positive wage premium is given to public sector employees.

7. Conclusion

In this paper, we apply the instrumental quantile regression estimator of Chernozhukov and Hansen (2004b and 2005b) to data from the German Socio Economic Panel to examine the wage structure in the public and private sector in West Germany. Since the original estimator proposed by Chernozhukov and Hansen does not allow for interaction terms between the endogenous variable and the covariates if we do not have additional instruments, we propose two estimators that allow for this possibility. We derive their asymptotic distributions and compare their small sample behaviors by mean of a Monte-Carlo simulation.

The rich data set gives us sensible instruments related to the occupational status of the employees' parents, in particular we know if the father of each employee was a civil servant at the time the employee was 16 years old. By applying the instrumental quantile regression estimator we can correct for endogenous sector choice and provide a full characterization of the effect of the public sector status on the distribution of wages. Furthermore, the estimators proposed in this paper allow different returns to human capital in both sectors. Different tests show that it is important to allow for all these three possibilities: there is endogenous sector choice, the public sector wage premium is different at different parts of the distribution and the returns to education and experience are different in the public sector.

The results assuming exogenous sector choice give a negative mean public sector wage premium and show that the wage distribution is more compressed in the public sector. Correcting for endogeneity reverses the findings concerning the mean premium but preserves the more compressed structure of the public sector earnings distribution. The Roy model (Roy 1951 and Heckman and Honore 1990) predicts that individuals will be positively selected towards the sector with higher wage inequality. This prediction is in fact confirmed by the data since we find positive selection into the private sector. Allowing different returns to human capital in the two sectors, we find that the public sector also reduces the between-group inequality by giving smaller returns to education. Thus, it seems that there is a political pressure on the government not to pay low wages to its less skilled employees and not to pay very high wages to its most skilled employees[13]. Finally, returns to experience remain positive almost until the end of the career in the public sector, probably as a consequence of the rigid hierarchical pay structure and the automatic salary increase with seniority.

[13] This is true for observed and unobserved (for the econometrician) skills.

Acknowledgements

I thank Michael Lechner, Patrick Puhani and seminar participants at the University of St. Gallen, the University of Darmstadt, the 2005 annual meeting of the Swiss Society of Economics and Statistics at Zürich, the second world conference of SOLE and EALE at San Francisco.

References

Abadie, A., 1995, Changes in Spanish labor income structure during the 1980s: a quantile regression approach. CEMPFI Working Paper No. 9521.

Abadie, A., Angrist, J. and G. Imbens, 2002, Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. Econometrica 70, 91-117.

Ahn, H. and J.L. Powell, 1993, Semiparametric estimation of censored selection models with a nonparametric selection mechanism. Journal of Econometrics 58, 3-29.

Angrist, D.J., Imbens, G.W. ad D.B. Rubin, 1996, Identification of causal effects using instrumental variables. Journal of the American Statistical Association 91, 444-455.

Amemiya T., 1982, Two Stage Least Absolute Deviations Estimators. Econometrica 50, 689-711.

Andrews, D.W.K. and M. Schafgans, 1996, Semiparametric estimation of the intercept of a sample selection model. Review of Economic Studies 65, 497-517.

Blinder, A., 1973, Wage discrimination: reduced form and structural estimates. Journal of Human Resources 8, 436-455.

Bofinger, E., 1975, Estimation of a density function using order statistics. Australian Journal of Statistics 17, 1-7.

Buchinsky, M., 1991, The Theory and Practice of Quantile Regression. Ph. D. Dissertation, Harvard University.

Buchinsky, M., 1998, The dynamics of changes in female wage distribution in the US: a quantile regression approach. Journal of Applied Econometrics 13, 1-30.

Buchinsky, M., 2001, Quantile regression with sample selection: estimating women's return to education in the US. Empirical Economics 26, 87-113.

Chamberlain, G., 1994, Quantile regression, censoring and the structure of wages, in: C. Sims (Ed.), Advances in Econometrics, Elsevier, New York, pp. 171-209.

Chen, L. and S. Portnoy, 1996, Two-stage regression quantiles and two-stage trimmed least-squares estimators for structural equation models. Communication in Statistics: Theory and Methods 25, 1005-1032.

Chen, X., Linton, O. and I. Van Keilegom, 2003, Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71, 1591-1608.

Chernozhukov, V. and C. Hansen, 2001, An IV Model of Quantile Treatment Effects. MIT Working Paper.

Chernozhukov, V. and C. Hansen, 2004a, The impact of 401K participation on savings: an IV-QR analysis. Review of Economics and statistics 86, 735-751.

Chernozhukov, V. and C. Hansen, 2004b, Instrumental quantile regression. MIT Working Paper.

Chernozhukov, V. and C. Hansen, 2005a, An IV Model of Quantile Treatment Effects. Econometrica 73, 245-262.

Chernozhukov, V. and C. Hansen, 2005b, Instrumental quantile regression inference for structural and treatment effect models. Journal of Econometrics, forthcoming.

Chesher, A., 2003, Identification in nonseparable Models. Econometrica 71, 1405-1441.

Chesher, A., 2005, Nonparametric identification under discrete variation. Econometrica forthcoming.

Dustmann, C. and C., Van Soest, 1998, Public and private sector wages of male workers in Germany. European Economic Review 42, 1417-41.

Goddeeris, J.H., 1988, Compensating differentials and self-selection : an application to lawyers. Journal of Political Economics 96, 411-428.

Gunderson, M., 1979, Earnings differentials between the public and private sectors. Canadian Journal of Economics 12, 228-242.

Gyourko, J. and J. Tracy, 1988, An analysis of public- and private- sector wages allowing for endogenous choices of both government and union status. Journal of Labor Economics 6, 229-253.

Heckman, J., 1990, Varieties of sample selection bias. American Economic Review 80, 313-318.

Heckman, J. and B. Honore, 1990, The empirical content of the Roy model. Econometrica 58, 1121-1149.

Hong, H. and E. Tamer, 2003, Inference in censored models with endogenous regressors. Econometrica 71, 905-932.

Honore, B. and L. Hu, 2004, On the performance of some robust instrumental variables estimator. Journal of Business and Economic Statistics 22, 30-39.

Horowitz, J.L., 1992, A smoothed maximum score estimator for the binary response model. Econometrica 60, 505-531.

Horowitz, J.L., 2002, Bootstrap critical values for tests based on the smoothed maximum score estimator. Journal of Econometrics 11, 141-167.

Ichimura H., 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics 58, 71-120.

Klein, R. and R. Spady, 1993, An efficient semiparametric estimator of the binary response model. Econometrica 61, 387-421.

Kim, J., and D. Pollard, 1990, Cube root asymptotics. Annals of Statistics 18, 191-219.

Koenker, R. and G. Bassett, 1978, Regression Quantiles. Econometrica 46, 33-50.

Koenker, R. and V. D'Orey, 1987, Computing Regression Quantiles. Applied Statistics 36, 383-393.

Kordas, G., 2005, Smoothed binary regression quantiles. Journal of applied econometrics, forthcoming.

Lee, S., 2004, Endogeneity in quantile regression models: a control function approach. CEMMAP working paper 08/04.

Ma, L. and R. Koenker, 2004, Quantile regression methods for recursive structural equation models. Journal of Econometrics, forthcoming.

Manski, C.F., 1975, Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3, 205-228.

Manski, C.F., 1985, Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. Journal of Econometrics 32, 65-108.

Melly, B., 2005a, Public-private sector wage differentials in Germany: evidence from quantile regression. Empirical Economics 30, 505-520.

Melly, B., 2005b, Decomposition of differences in distribution using quantile regression: asymptotic results. Mimeo.

Mueller, R., 1998, Public-private sector wage differentials in Canada: evidence from quantile regressions. Economics Letters 60, 229-235.

Newey, W., 1984, A method of moments interpretation of sequential estimators. Economics Letters 14, 201-206.

Newey, W. and D.L. McFadden, 1994, Large sample estimation and hypothesis testing, in: R.F. Engle and D.L. McFadden (Eds.), Handbook of Econometrics, Vol. IV Elsevier Science, Amsterdam.

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

Portnoy, S. and R. Koenker, 1997, The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. Statistical Science 12, 279-300.

Poterba, J. and K., Rueben, 1995, The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods. NBER Working Paper No. 4734

Powell, J.L., 1983, The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators. Econometrica 51, 1569-1575.

Powell, J.L., 1984, Least absolute deviations estimation for the censored regression model. Journal of Econometrics 25, 303-325.

Roy, A.D., 1951, Some thoughts on the distribution of earnings. Oxford Economic Papers 3, 135-146.

Sekhon, J.S. and W.R. Mebane, 1998, Genetic optimization using derivatives: theory and application to nonlinear models. Political Analysis 7, 189-213.

Silverman, B.W., 1986, Density estimation. Chapman and Hall, London.

Smith, S., 1976, Pay differentials between federal government and private sectors workers. Industrial and Labour Relations Review 29, 233257.

SOEP Group, 2001, The German Socio-Economic Panel (GSOEP) after more than 15 years – Overview, in: E. Holst, D.R. Lillard and T.A. DiPrete, (Eds.), Proceedings of the 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000), Vierteljahrshefte zur Wirtschaftsforschung (Quarterly Journal of Economic Research), 70:1, pp. 7-14.

Strotz, R. and H. Wold, 1960, A triptych on causal systems. Econometrica 28, 417-463.

Appendix A: Instrumental Variable Quantile Regression

Chernozhukov and Hansen (2005a) focus on modeling and nonparametric identification and show how the results can be derived from primitive conditions. The estimator for a single quantile is defined and studied in Chernozhukov and Hansen (2004b). The properties of the instrumental variable quantile regression process and of the inference process and test statistics derived from it are established in Chernozhukov and Hansen (2005b). An application can be found in Chernozhukov and Hansen (2004a).

Chernozhukov and Hansen (2004b and 2005b) note that

$$\Pr\left(Y \le D'\alpha(\theta) + X'\beta(\theta) \middle| X, Z\right) = \theta$$

is equivalent to the statement that 0 is the θ^{th} quantile of $Y - D'\alpha(\theta) - X'\beta(\theta)$ conditional on (X, Z). Thus, the problem is to find parameters such that

$$0 = \arg\min_{\gamma} E\left[\rho_\theta\left(Y - D'\alpha(\theta) - X'\beta(\theta) - \gamma Z\right)\right]. \tag{A.1}$$

The finite-sample analog of this procedure is simple and implies only estimation of traditional quantile regression along a $\dim(\alpha)$-dimensional grid. In the simplest case where *D* and *Z* are one-dimensional (1 instrument and 1 endogenous variable), the procedure consists simply in finding α such that the traditional quantile regression of $Y - D'\alpha$ on Z and X gives a coefficient of zero on *Z*. In order to formalize the estimator in the general case, allowing for estimated weights and instruments, define the quantile regression objective function:

$$Q_n(\theta, \alpha, \beta, \gamma) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\rho_\theta\left(Y_i - D_i'\alpha - X_i'\beta - \hat{\Phi}_i(\theta)'\gamma\right)\hat{V}_i(\theta)\right]$$

where $\Phi_i(\theta) \equiv \Phi(\theta, X_i, Z_i)$ is an *r*-vector of instruments and $V_i(\theta) \equiv V(\theta, X_i, Z_i) > 0$ is a weight function. $\hat{\Phi}_i(\theta)$ and $\hat{V}_i(\theta)$ are consistent estimates of $\Phi_i(\theta)$ and $V_i(\theta)$. The estimation procedure is defined as follows:

$\hat{\alpha}(\theta) = \arg\inf_{\alpha \in A} W_n(\theta,\alpha)$, $W_n(\theta,\alpha) := n\hat{\gamma}(\theta,\alpha)'\hat{\mathrm{A}}(\theta,\alpha)\hat{\gamma}(\theta,\alpha)$ such that

$\left(\hat{\beta}(\theta,\alpha),\hat{\gamma}(\theta,\alpha)\right) = \arg\inf_{(\beta,\gamma)} Q_n(\theta,\alpha,\beta,\gamma)$, so that

$\left(\hat{\alpha}(\theta),\hat{\beta}(\theta)\right) = \left(\hat{\alpha}(\theta),\hat{\beta}(\theta,\hat{\alpha}(\theta))\right)$.

$\hat{\mathrm{A}}(\theta,\alpha) = \mathrm{A}(\theta,\alpha) + o_p(1)$ and $\mathrm{A}(\theta,\alpha)$ is positive definite uniformly in $\alpha \in A$. It is convenient to set $\mathrm{A}(\theta,\alpha)$ equal to the inverse of the asymptotic covariance matrix of $\sqrt{n}\left(\hat{\gamma}(\theta,\alpha) - \gamma(\theta,\alpha)\right)$. In this case, more weights are given to the instruments that are more precisely estimated and $W_n(\theta,\alpha)$ is the Wald statistics for testing $\gamma(\theta,\alpha) = 0$.

Under some technical regularity conditions, Chernozhukov and Hansen (2004b) derives the asymptotic distribution of $\left(\hat{\alpha}(\theta),\hat{\beta}(\theta)\right)$:

$$\sqrt{N}\begin{pmatrix} \hat{\alpha}(\theta) - \alpha(\theta) \\ \hat{\beta}(\theta) - \beta(\theta) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Lambda(\theta) = (K',L')'S(K',L')\right) \tag{A.2}$$

where, for $\Psi = V \cdot [X',\Phi']'$ and $\varepsilon = Y - D'\alpha(\theta) - X'\beta(\theta)$, $S = \theta(1-\theta)E[\Psi\Psi']$,

$K = (J_\alpha' H J_\alpha)^{-1} J_\alpha' H$, $H = \bar{J}_\gamma' A(\theta,\alpha)\bar{J}_\gamma$, $L = \bar{J}_\beta M$, $M = I_{k+r} - J_\alpha K$,

$J_\alpha = E\left[f_\in\left(0|X,Z,D\right)\Psi D'\right]$, and $\left[\bar{J}_\beta',\bar{J}_\gamma'\right]'$ is a partition of $E\left[f_\varepsilon\left(0|X,Z\right)\Psi\Psi'/V\right]^{-1}$ such that $\bar{J}_\beta$ is a $k \times (k+l)$ matrix and $\bar{J}_\gamma$ is a $l \times (k+l)$ matrix.

Efficiency can be achieved by choosing $V^* = f_\varepsilon\left(0|X,Z\right)$ and $\Phi^* = E\left[D\upsilon^*|X,Z\right]/V^*$, where $\upsilon^* = f_\varepsilon\left(0|D,X,Z\right)$. Then, the asymptotic variance simplifies to $\theta(1-\theta)E[\Psi\Psi']^{-1}$ and attains the efficiency bound.

Appendix B: Asymptotic distribution of the estimator (SIQR)

The notation necessary to describe the asymptotic distribution of the estimator described in section 3.1 which combines the sample selection correction of Buchinsky (1998) and the instrumental variable quantile regression estimator of Chernozhukov and Hansen is complicate since it is as a 3-steps estimator and the asymptotic variance of the first-step estimate appears in the second step and the asymptotic variance of the second step estimate appears in the third step. However, the procedure is intuitively straightforward:

1st step: We regress D on X and Z and we obtain the estimated coefficients $\hat{\alpha}_X$ and $\hat{\alpha}_Z$ and their asymptotic covariance matrix Λ_α. This step can be estimated using different existing parametric (logit, probit) or semiparametric (Klein and Spaddy (1993), Ichimura (1993)) estimators[14].

2nd step: Denote by $\hat{g} = X\hat{\alpha}_X + Z\hat{\alpha}_Z$ the estimated index and by $P_S(\hat{g}) = (P_{S1}(\hat{g}),...,P_{SS}(\hat{g}))$ a polynomial vector in $\hat{g}$ of order S. We estimate now the quantile regression of Y on X and $P_S(\hat{g})$ separately for observations with $D=0$ and $D=1$:

$$Y_i = X_i\hat{\beta}_0(\theta) + P_S(\hat{g}_i)\hat{\kappa}_0(\theta) + \hat{\varepsilon}_{0i}(\theta),\ \{i : D_i = 0\}$$

$$Y_i = X_i\hat{\beta}_1(\theta) + P_S(\hat{g}_i)\hat{\kappa}_1(\theta) + \hat{\varepsilon}_{1i}(\theta),\ \{i : D_i = 1\}.$$

The bias terms $h_{\theta,0}(g)$ and $h_{\theta,1}(g)$ are approximated by $P_S(\hat{g}_i)\hat{\kappa}_0(\theta)$ and $P_S(\hat{g}_i)\hat{\kappa}_1(\theta)$. The asymptotic distributions of $\hat{\beta}_1(\theta)$ can be derived directly following Buchinsky (1998):

$$\sqrt{n}\left(\hat{\beta}_1(\theta) - \beta_1(\theta)\right) \to N\left(0, \Lambda_{\beta_1(\theta)}\right)$$

[14] In principle, the results of Chen, Linton and Van Keilegom (2003) allow to use nonparametric first step estimators if we use high-order kernels and undersmoothing. Thus an estimator in the spirit of Ahn and Powell (1993) could be built for quantile regression. Note however that the single index assumption must be maintained in the second step and the finite sample properties of this estimator should be pretty bad, particularly with the sample size that we have in our application.

where $\Lambda_{\beta_1(\theta)}$ is the $k-1\times k-1$ top-left submatrix of

$$\Delta_{fr1}^{-1}\left(\theta(1-\theta)\Delta_{rr1}+\Delta_{frx1}\Lambda_{\alpha}\Delta_{frx1}^{T}\right)\Delta_{fr1}^{-1}$$

where $\Delta_{fr1}=E\left[f_{\varepsilon_1(\theta)}\left(0|r_1\right)r_1r_1'\right]$, $\Delta_{frx1}=E\left[f_{\varepsilon_1(\theta)}\left(0|r_1\right)\frac{dh_{\theta,1}(g)}{dg}'\kappa_1(\theta)r_1X\right]$, $\Delta_{rr1}=E\left[r_1r_1'\right]$

and $r_1'=D\cdot\left(X,h_{\theta,1}(g)\right)$.

Similarly, the asymptotic distribution of $\hat{\beta}_0(\theta)$ is given by

$$\sqrt{n}\left(\hat{\beta}_0(\theta)-\beta_0(\theta)\right)\rightarrow N\left(0,\Lambda_{\beta_0(\theta)}\right)$$

where $\Lambda_{\beta_0(\theta)}$ is the $k-1\times k-1$ top-left submatrix of

$$\Delta_{fr0}^{-1}\left(\theta(1-\theta)\Delta_{rr0}+\Delta_{frx0}\Lambda_{\alpha}\Delta_{frx0}^{T}\right)\Delta_{fr0}^{-1}$$

where $\Delta_{fr0}=E\left[f_{\varepsilon_0(\theta)}\left(0|r_0\right)r_0r_0'\right]$, $\Delta_{frx0}=E\left[f_{\varepsilon_0(\theta)}\left(0|r_1\right)\frac{dh_{\theta,0}(g)}{dg}'\kappa_0(\theta)r_0X\right]$, $\Delta_{rr0}=E\left[r_0r_0'\right]$

and $r_0'=(1-D)\cdot\left(X,h_{\theta,0}(g)\right)$.

The asymptotic covariance between $\hat{\beta}_0(\theta)$ and $\hat{\beta}_1(\theta)$ is given by the $k-1\times k-1$ top-left submatrix of

$$\Delta_{fr0}^{-1}\Delta_{frx0}\Lambda_{\alpha}\Delta_{frx1}^{T}\Delta_{fr1}^{-1}.$$

We note that $\hat{\beta}_0(\theta)$ and $\hat{\beta}_1(\theta)$ are correlated although they use different sets of observations since they use the same first step estimate of $\Pr\left(D=1|X,Z\right)$. We define Λ_{β} to be the whole variance-covariance matrix of $\hat{\beta}_0(\theta)$ and $\hat{\beta}_1(\theta)$.

3rd step: We use the Chernozhukov and Hansen (2004b) estimator with $Y-(1-D)X\hat{\beta}_0-DX\hat{\beta}_1$ as dependent variable and only a constant as exogenous regressor. The new weighted quantile regression objective function is given by

$$Q_n(\theta,\alpha,\delta,\gamma) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\rho_\theta\left(Y_i - \alpha - D_i'\delta - (1-D_i)X_i'\hat{\beta}_0 - D_i X_i'\hat{\beta}_1 - \hat{\Phi}_i(\theta)'\gamma\right)\hat{V}_i(\theta)\right].$$

The estimation procedure is defined as follows:

$$\hat{\delta}(\theta) = \arg\inf_{\delta\in A} n\hat{\gamma}(\theta,\delta)'\hat{A}(\theta,\delta)\hat{\gamma}(\theta,\delta), \text{ such that}$$

$$\left(\hat{\alpha}(\theta,\delta),\hat{\gamma}(\theta,\delta)\right) = \arg\inf_{(\alpha,\gamma)} Q_n(\theta,\alpha,\delta,\gamma), \text{ so that}$$

$$\left(\hat{\alpha}_0(\theta),\hat{\alpha}_1(\theta),\hat{\beta}_0(\theta),\hat{\beta}_1(\theta)\right) = \left(\hat{\alpha}\left(\theta,\hat{\delta}(\theta)\right),\hat{\alpha}\left(\theta,\hat{\delta}(\theta)\right)+\hat{\delta}(\theta),\hat{\beta}_0(\theta),\hat{\beta}_1(\theta)\right).$$

Here we consider only the exact identified case to keep the notation tractable but the overidentified case can be derived in a similar way by weighting the moment conditions.

The asymptotic distribution of $\hat{\alpha}_0(\theta)$ and $\hat{\delta}(\theta)$ can be derived applying the results for 2-steps GMM estimators (for instance Newey 1984)[15] since the true value $\alpha_0(\theta)$ and $\delta(\theta)$ solve the following moment conditions:

$$E\left[\left(1\left(Y_i < \alpha(\theta) - D'\delta(\theta) - (1-D)X'\beta_0(\theta) - DX'\beta_1(\theta)\right) - \theta\right)\left[1:\Phi(\theta)'\right]'V(\theta)\right] = 0.$$

The covariance matrix of these moment conditions is given by

$$S = \theta(1-\theta)E\left[\Psi(\theta)\Psi(\theta)'\right] \text{ where } \Psi(\theta) = V(\theta)\cdot\left[1,\Phi(\theta)'\right]'$$

and the derivatives of these moment conditions relative to α and δ are given by

$$J_\theta = E\left[f_\varepsilon\left(0\middle|X,\Phi(\theta),D\right)\Psi(\theta)[D,1]\right].$$

If $\beta_0(\theta)$ and $\beta_1(\theta)$ would be given and not estimated, the asymptotic variance of $\left(\hat{\alpha}_0(\theta),\hat{\delta}(\theta)\right)$ would be $J_\theta S J_\theta'$. However, the derivative of the moment condition relative to (β_0,β_1) is not zero but is given by

[15] Note that the results of Newey are valid only for smooth moment conditions. However, if we assume that the density of the error term at 0 is bounded away from zero and continuous, the moment conditions of the quantile regression estimators are asymptotically smooth and thus results of Powell (1984) or more generally Newey and Mc Fadden (1994, section 7) allows to apply the GMM framework.

$$J_\beta = \left[E\left[(1-D) V(\theta) f_\varepsilon \left(0 \middle| X, \Phi(\theta) \right) \Psi(\theta) X \right], E\left[D V(\theta) f_\varepsilon \left(0 \middle| X, \Phi(\theta) \right) \Psi(\theta) X \right] \right].$$

Therefore, following standard results for multi-step GMM estimators and since the moment condition of the second and third step are uncorrelated, the asymptotic distribution of $\left(\hat{\alpha}_0(\theta), \hat{\delta}(\theta) \right)$ is

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}(\theta) - \alpha(\theta) \\ \hat{\delta}(\theta) - \delta(\theta) \end{pmatrix} \to N\left(0, J_\theta^{-1} \left(S + J_\beta \Lambda_\beta J_\beta ' \right) J_\theta^{-1} ' \right).$$

The covariance matrix between $\left(\hat{\beta}_0(\theta), \hat{\beta}_1(\theta) \right)$ and $\left(\hat{\alpha}_0(\theta), \hat{\delta}(\theta) \right)$ is found to be equal to $\Lambda_\beta J_\beta ' J_\theta^{-1} '$. Finally, since the measures of inequality are a function of more than one quantile estimate, we need the covariance matrix of distinct quantile estimates. This is straightforward to derive using the results of Buchinsky (1998) and Chernozhukov and Hansen (2004b). Basically, the covariance matrix of the estimate at the quantiles θ_1 and θ_2 is the same as the covariance matrix of the estimate for a single quantile θ but with $\min(\theta_1, \theta_2) - \theta_1 \theta_2$ instead of $\theta(1-\theta)$ and all matrix that appears twice ($\Delta_{fr}, \Delta_{frx}$, J_β, J_θ) are evaluated once at θ_1 and once at θ_2.