

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

February 9, 2006

Abstract:

We apply the instrumental quantile regression estimator of Chernozhukov and Hansen (2004b and 2006) to examine the wage structure in the public and private sector in Germany. Assuming exogenous sector choice, we find 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. Since the original estimator loses its good properties if we allow the public sector premium to vary with the covariates, we propose computationally convenient estimators that achieve good small-sample properties. Applying these estimators, we find that returns to experience and education are generally higher in the private sector.

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
<http://www.siaw.unisg.ch/lechner/melly>

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. Thus, all taxpayers are concerned that government runs on an efficient basis. Second, public sector labor markets are different. There are a number of reasons, surveyed by Bender (1998), that earnings differential between the private and the public sector exists. Basically, the public sector is subject to political constraints and not to profit constraints: the goal of a politician is to be re-elected while the goal of a firm is to make profit (and, at least, no deficit). Therefore, issues such that pay equity and fairness can survive in the political market place more than in the economic market place and, hence, the whole distribution of wages is potentially different in the public sector.

Given these differences in the wage setting procedures and the possible consequences on the economic well-being of a country, 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. Early research (Smith 1976, Gunderson 1979) has used least-squares regressions to compare the predicted wages in both sectors conditionally on basic human capital characteristics. More recently, two different directions of research have complemented these first results. One strand of research concentrates on the mean public sector wage differential and takes account of possible non-random selection by Heckman (1979) /Lee (1978) correction for selectivity, endogenous switching regression models or fixed effect panel models.¹ They mostly find a significant selection bias but their results differ widely, probably because of weak (or absence of valid) instruments. A second strand of research complements the results for the mean by estimating the distribution of the public sector wage premium, mostly with quantile regression.² However, they must assume that the sector status is

¹ See e.g. Goddeeris (1988), Gyourko and Tracy (1988) and Dustmann and Van Soest (1998).

² See e.g. Poterba and Rueben (1995), Mueller (1998) and Melly (2005).

exogenous. The main conclusion from these studies is that the public sector compresses the conditional distribution of earnings.

The objective of this paper is to bring these two strands of research together. Until recently, it was impossible to allow simultaneously for endogenous sector choice and heterogeneous public sector gaps at different points of the distribution. During the last years, different instrumental variable methods for estimating endogenous quantile treatment effect have been proposed³. After comparing the different models and estimators, we use the proposal of Chernozhukov and Hansen (2004b and 2006) to estimate the effects of the public sector status on the entire wage distribution controlling for endogenous sector choice. We also present estimates obtained through the procedure of Abadie, Imbens and Angrist (2002) and find similar results. We use data from the German Socio-Economic Panel⁴, which contains background information on parents' economic status and provides us with reasonable instruments, used for instance by Dustmann and Van Soest (1998). 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. The estimator proposed by Chernozhukov and Hansen is a convenient estimator if the number of endogenous variables is small but becomes more and more difficult to implement when this number increases. In order to allow the public sector wage premium to vary with the covariates, we need to estimate a fully interacted model which would be practically infeasible in term of computation time. Therefore, we propose two estimators that are tractable in this configuration. Their asymptotic distributions are derived and a Monte-Carlo simulation shows that they have good finite-sample properties.

³ Abadie, Angrist and Imbens (2002), Chernozhukov and Hansen (2004b and 2006), Chesher (2003 and 2005), Ma and Koenker (2006), Lee (2004), Honore and Hu (2004), Hong and Tamer (2003), among others.

⁴ For an English language description of the GSOEP see SOEP Group (2001).

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 controls for endogenous sector choice and analyzes the public and private sector wage distributions. Heterogeneous public sector gaps and endogenous sector choice are shown to be really present and important. In summary, correcting for endogenous sector choice reverses the findings concerning the mean premium but preserves the more compressed structure of the public sector earnings distribution. Applying the new estimators, we find that returns to education are higher and that the experience-wage profile is more concave in the private sector.

Section 2 compares the different estimators 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 2006). Section 3 proposes two extensions allowing for regressors fully interacted with the endogenous variable and gives the results of a Monte-Carlo simulation. 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 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:

$$Y = X\beta(\tau) + \varepsilon \text{ and } F_{\varepsilon}^{-1}(\tau|X) = 0,$$

where $F_{\varepsilon}^{-1}(\tau|X)$ denotes the τ^{th} quantile of ε conditionally on X . Koenker and Bassett (1978) propose to estimate the τ^{th} regression quantile by solving

$$\hat{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^K} \sum_{i=1}^N \rho_{\tau}(y_i - x_i' \beta),$$

where ρ_{τ} is the *check function*, $\rho_{\tau}(z) = z(\tau - 1(z \leq 0))$ and $1(\cdot)$ is the indicator function.

They show $\sqrt{N}$ consistency and asymptotic normality of $\hat{\beta}(\tau)$. Buchinsky (1991) shows that this estimator has a GMM interpretation since the true $\beta(\tau)$ satisfies the following moment condition

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

Increasing τ continuously from 0 to 1, we can trace the entire distribution of Y conditionally on X . By complementing a model of conditional central tendency with a family of models for conditional quantiles, we are able to achieve a 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 by 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 “fitted value” approach is not consistent when the quantile treatment effect differs across quantiles which is, however, the main motivation for using quantile regression.

Other approaches have been considered, as well. Chesher (2003) develops a general nonparametric 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 (2006) 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. Although Chesher (2005) shows that extensions for discrete variables are possible, he excludes the binary endogenous variable case, which is the situation encountered in the public-private sector application.

Abadie, Imbens and Angrist (2002) propose a parametric estimator based on the LATE model of Imbens and Angrist (1994). Their estimator apply only to a special case: a binary treatment variable D and a binary instrument Z .⁵ They impose an independence condition (Z is independent of the errors in the outcome and selection equations) and a monotonicity condition (the direction of the effect of Z on the participation decision is the same for all individuals). They show that under these assumptions the marginal distributions of the potential outcome are identified for the sub-population of compliers and suggest an ingenious estimator that can be interpreted as a re-weighted standard quantile regression estimator.

If we assume that the quantile treatment effects are homogeneous, a more direct estimation strategy uses the exclusion restrictions (instruments) directly in the GMM framework. Suppose we have a structural relationship defined by

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

$$\tau \rightarrow D' \alpha(\tau) + X' \beta(\tau) \text{ is strictly increasing in } \tau, \quad (2)$$

$$D = \delta(X, Z, V). \quad (3)$$

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

$$\Pr(Y \leq D'\alpha(\tau) + X'\beta(\tau) | X, Z) = \tau, \quad (4)$$

thus providing the moment conditions

$$E\left[\left(\tau - 1(Y \leq D'\alpha(\tau) + X'\beta(\tau))\right)(X', Z')'\right] = 0. \quad (5)$$

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,⁶ we could estimate $\alpha(\tau)$ and $\beta(\tau)$ by traditional GMM. This strategy was used by Hong and Tamer (2003), Chen, Linton and Van Keilegom (2003) and Honore and Hu (2004) to construct estimators.⁷ 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 $\mathbb{R}^{\dim(\alpha) + \dim(\beta)}$,

⁵ Extensions allowing for continuous instruments in a similar way as the local instrumental variable estimator for the mean (Heckman and Vytlačil, 1999) are possible. See Carneiro and Lee (2005) for recent developments.

⁶ See for instance assumption R2 in Chernozhukov and Hansen (2006). Basically, it requires that a density-weighted covariance matrix between D and Z is of full rank.

⁷ Honore and Hu use the moment condition directly. Chen, Linton and Van Keilegom and Hong and Tamer use a minimum distance framework with nonparametric first step estimation. Chen, Linton and Van Keilegom propose an extension for partially linear models and Hong and Tamer for censored models.

thus rendering the application of these estimators almost impossible in data sets typically found in microeconometrics⁸.

The instrumental variable quantile regression (IVQR) 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 $\alpha(\tau)$, we could estimate $\beta(\tau)$ consistently by regressing $Y - D'\alpha(\tau)$ on X with traditional quantile regression. In reality, we don't know $\alpha(\tau)$ but we have instruments Z . Thus, we can try different values for $\alpha(\tau)$ and regress $Y - D'\alpha(\tau)$ on X and Z . If the model is identified, the true value of $\alpha(\tau)$ is the only one for which the coefficients on Z are zero. This reduces considerably the computation time since we need to perform a grid search only on $\dim(\alpha)$ which is frequently small. Moreover, quantile regression can be solved very fast using interior point algorithms (Portnoy and Koenker 1997).⁹

In the public-private sector application, we have an endogenous dummy variable. Thus, only the approaches of Abadie, Angrist and Imbens (2002) and Chernozhukov and Hansen (2004b) can be applied. The comparison of both models shows that Abadie, Angrist and Imbens (2002) impose more restrictive conditions on the choice equation but allows for heterogeneity¹⁰ in responses. When the assumptions of both models are satisfied, they estimate the same quantity and, therefore, a comparison of both sets of estimates provides a useful robustness check.

We will concentrate principally on the approach of Chernozhukov and Hansen for two reasons. First, we have in our application five binary instruments and a relative small sample

⁸ 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.

⁹ Details about the procedure and the asymptotic distribution of the estimator are given in Appendix A.

¹⁰ Here we mean heterogeneity conditional on X and the quantile of interest. Naturally, both estimators allow for different treatment effect at different quantiles.

such that it is really difficult to estimate heterogeneous quantile treatment effects. The estimator of Chernozhukov and Hansen (2004b) allows us to use the five instruments to estimate the quantile treatment effects. As a comparison, we will also apply the estimator of Abadie, Angrist and Imbens (2002) with only one - the most significant - instrument. Second, only Chernozhukov and Hansen (2005b) 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 instrumental variable quantile regression but are found using the whole instrumental variable quantile regression process.

3. Estimation of interacted models

In order to allow the public sector wage premium to vary with X , we need to estimate a fully interacted model. This increases dramatically the number of endogenous covariates to $\dim(D)\dim(X)$, which is equal to 8 in our application. Although the moment conditions will be highly correlated, we can generate a sufficient number of instruments by interacting the instruments with the exogenous regressors. However, the computation time needed to perform grid searches over such high-dimensional parameter spaces renders the procedure infeasible. Therefore two computationally tractable estimators for the estimation of fully interacted models are proposed in this section. 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 . In a second step, we estimate the global parameters by 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 the series estimator

suggested by Newey (1988). 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 with selectivity.

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:

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

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

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

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 $\Pr(D=1) \rightarrow 1$ and others with $\Pr(D=1) \rightarrow 0$. 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. The estimation often rests on just a handful of observations surpassing the growing threshold which may be hard to distinguish from unreasonable outliers.

Thus, we propose to estimate only the slope coefficients with the sample selection procedure of Buchinsky to obtain $\sqrt{N}$ consistent and asymptotically normally distributed estimates $\hat{\beta}_0(\tau)$ and $\hat{\beta}_1(\tau)$. 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 $\hat{\beta}_0(\tau)$, $\hat{\beta}_1(\tau)$, $\hat{\alpha}_0(\tau)$ and $\hat{\alpha}_1(\tau)$ are $\sqrt{N}$ consistent and asymptotically jointly normally distributed. All the details about the estimator (called SIVQR), the asymptotic distribution and the proofs can be found in Appendix B.

3.2. Integration of nonparametric first step estimates

When the distribution of the covariate vector X has finite support, 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 IV quantile regressions 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. The asymptotic distribution of this estimator (called MDIVQR) can be found in Appendix C. To 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 does not require the single index assumption for consistency.

On the other hand, if some of the exogenous variables are continuous, the problem becomes more complicated. We can first estimate the quantile function 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 finite-dimensional parameters. 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 subsection shows that the small sample properties of the minimum distance estimator are worse than those of the SIVQR. Second, we would lose the good computational property of the estimator which is the principal reason to use this estimator.

3.3. *Finite sample properties of these estimators*

In order to compare the performance of the two proposed estimators with the IVQR and to evaluate the costs in term of variance of allowing for interaction terms, we present the results of a Monte- Carlo simulation. We choose the simplest data generating process that allows us to consider these questions:

$$\begin{aligned}
Y &= \alpha + X\beta_x + D\beta_w + XD\beta_{xw} + (1 + 0.2D)U \\
X &\sim b(0.5); Z \sim N(0,1); D \sim 1(0.5 - X + Z + \varepsilon < 0) \\
\varepsilon &\sim t_3; U \sim t_3; \text{Cov}(\varepsilon, U) = 0.8.
\end{aligned} \tag{6}$$

We have only one endogenous binary variable, one exogenous binary variable and one instrument. The amount of endogeneity is quite high with a correlation of 0.8 between both error terms. Traditional parametric models for the binary choice model are not consistent since the error term is t -distributed with 3 degrees of freedom. The treatment effect varies with the quantile because of heteroscedasticity. 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, 400 and 1600 and draw 10000 replications. In each replication, we apply the traditional quantile regression estimator (QR), the IVQR of Chernozhukov and Hansen with and without interaction term, the SIVQR and finally the MDIVQR. The Chernozhukov and Hansen estimators are obtained by searching on regular

grids with steps of length 0.01. Since we have only one endogenous and one exogenous regressors, the grid search is only 2-dimensional if we allow for an interaction term. In the application, we have 8 exogenous regressors and grid search becomes infeasible. For the SIVQR the selection equation is estimated by the Klein and Spady (1993) semiparametric estimator. Since the results are similar for different quantiles, we present only the outputs for the median regression. Tables 1 and 2 give the bias, the standard error (S.E.) and the mean squared error (MSE) for each estimator in the case without and with interaction term respectively.

These simulations confirm that the 3 estimators allowing for an interaction term between D and X are consistent in both cases while the IVQR not allowing for an interaction term converges to values difficult to interpret if there is an interaction term. $\sqrt{N}$ convergence rate is also validated since quadrupling the sample size divides the standard errors by about 2 and the MSE by about 4. We note also that QR is heavily biased, as a consequence of the high level of endogeneity presents in the data generating processes. Naturally, allowing for endogeneity and interaction terms has a price in term of standard errors: the standard errors of the IVQR is up to 80% higher than those of QR and allowing for an interaction term can double the standard errors of the estimates. Now, if we compare the 3 estimators allowing for an interaction term and endogeneity, we remark that the SIVQR is always more precise than the 2 others and that the differences are not negligible (at least 20% and up to 100%). The MDIVQR is asymptotically equivalent with the IVQR allowing for an interaction term but has better small sample properties. Thus, based on the computational and statistical properties of the estimators and the results of these Monte Carlo simulations, we apply the SIVQR estimator in Section 6.4.

4. Decomposition of differences in distribution

The most basic approach to explore wage differentials between groups or sectors involves estimating an earnings regression using pooled data 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 differences at the mean. It allows very easily to decompose the total difference into a part explained by different characteristics and a part explained by coefficients. 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 (2006) 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 $\hat{\beta} = (\hat{\beta}(\tau_1), \dots, \hat{\beta}(\tau_j), \dots, \hat{\beta}(\tau_J))$ be the quantile regression coefficients estimated at J different quantiles $0 < \tau_j < 1$, $j = 1, \dots, J$. Integrating over all quantiles and over all observations, a natural estimator of the θ^{th} unconditional quantile of the dependent variable is given by

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

Melly (2006) 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:

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

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: effect of endogenous sector choice, effect of differences in coefficients and effect 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) 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 substantial 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.¹¹ As the sample includes only wage earners, the results must be interpreted conditional on the selected sample. However since we concentrate on males, we expect that this selection bias is not important. Finally, all observations with a missing value for one of the variables have been excluded.¹² The final dataset contains 3125 observations.

Table 3 defines the variables we use for our empirical analyses. Y , the dependent variable, is $Lnghwage$, the logged gross hourly wage. X , the vector of regressors assumed to be exogenous, contains a quadratic in potential experience and 5 educational dummies. 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 distinguish between

¹¹ As a sensitivity check, we have also repeated the estimation procedure (except the bootstraps) only with men between 30 and 50 years and we have found no noteworthy difference.

¹² 245 observations have been excluded because of missing values.

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 the bandwidth was chosen according Silverman (1986) rule of thumb. 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 10th 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 90th 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 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, identification requires exclusion restrictions. In many studies, the data is not rich enough to provide appropriate instruments and identification assumptions are sometimes doubtful. 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 to 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, this probability is only of 21%.

6. Empirical results

6.1. *Exogenous sector choice*

As a benchmark, we first estimate the public-private sector wage differential assuming that the sector choice is exogenous. The first method used is the simple dummy variable approach. We regress 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 variances¹³. The densities were estimated by the method of Powell (1984) using a normal kernel and a bandwidth following the Bofinger (1975) rule. The estimated coefficients for the median regressions 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.

¹³ 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 (2004b) for the IVQR and of Melly (2006) for the decomposition procedure.

These results are correct only if the returns to individual characteristics are the same in both sectors. In order to test this restriction we estimate a fully interacted model where all characteristics are interacted with the public sector dummy and we test 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. 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 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 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 (2005). However, we should recall that these results were obtained assuming exogenous sector choice. This unlikely to be satisfied assumption will be abandoned 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 and, since the logit depends for consistency heavily on the distributional assumption, by the semiparametric estimator of Klein and Spady (1993). We implement the Klein and Spady estimator as in Gerfin (1996).¹⁴

The estimated coefficient vectors are normalized such that the slope coefficients have norm 1.

The results of the logit and of the Klein and Spady estimations are given in Table 6. The estimated coefficients are not fundamentally different. The standard errors of the logit estimates are generally lower than those of the semiparametric estimates, as expected, but the differences are not huge. The probability of working in the public sector increases 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 and find at least two significant instruments: $Fcivil$ is significantly different from zero at the 1 per mil and $Mnwork$ is significant at the 5 % level. If we believe the logit estimates, $Fwhite$ is significant at the 10% level. The Wald test for testing the hypothesis that the coefficients of all 5 instruments are equal to zero gives a value of 19 with the logit and 23 with the Klein and Spady estimator. We can thus reject the null hypothesis at all sensible significance levels.

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 2006) described in Appendix A. Given the relative small sample size, we do not try to weight the observations and choose $A(\theta, \alpha)$ to be the

¹⁴ A bandwidth of 0.3 was chosen. Note that the estimates of the sector choice equation are not really sensitive to the choice of the bandwidth and that the second step estimates (wage equations corrected for selection) are even less sensitive.

inverse of the asymptotic covariance matrix of $\sqrt{n}(\hat{\gamma}(\theta, \alpha) - \gamma(\theta, \alpha))$. The parameter space for α is 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 use equally spaced grids 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 endogenous sector choice inverts the conclusion: the majority of public sector employees is over- and not underpaid. Thus, there is positive selection into the private sector and negative into the public sector. The direction of the selection effect is correctly predicted by the Roy (1951, see also the discussion in Heckman and Honore 1990) model. Employees with an absolute disadvantage have a comparative advantage in the sector in which earnings are more concentrated. Thus, individuals will be positively selected towards the sector with higher wage inequality, the private sector.

The evolution of the premium when we let θ vary between 0 and 1 is similar whether we control for endogeneity or not. 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 to 1.5 if we control for endogenous sector choice. However, given the variances of the estimates, these extreme results should be kept with caution. In any cases, the different distributions of wages in both sectors are not caused by different distributions of unobserved ability.

It is not possible to draw significant conclusions for a single quantile because the standard errors of the estimated coefficients are almost 10 times higher if we allow for endogeneity. As explained in Appendix A, an efficient estimator can be obtained by using estimated optimal instruments and weights. We implemented the optimal estimator but were disappointed

because the standard errors of the estimates do not really decreased. However, we can obtain significant results by considering the whole instrumental quantile regression process. Chernozhukov and Hansen (2006) propose inference procedures to evaluate the impact of the treatment on the entire distribution of outcomes. They suggest a resampling 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, a difficult task requiring the quite arbitrary choice of the bandwidths. We use the Smirnov-Cramer-Von-Misses statistic and estimate Anderson-Darling weights by resampling. We estimate the critical values by constructing 1000 replications with 250, 2000 and 3125 observations drawn with replacement¹⁵ and estimate the instrumental quantile regression process on $\tau \in [0.1, 0.9]$.

The p -values for 5 hypotheses are given in Table 7. In general, the larger the subsample size the more conservative are the tests but they give all the same results for a confidence level of 5%. 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 the hypothesis of exogeneity, confirming the need to instruments for the sector choice. This confirms the visual impression of Figures 2 and 4 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. Thus, heterogeneous public sector wage gaps and endogenous sector choice are significant at the 5% significant level.

¹⁵ We want to check for the sensitivity of the results to the subsample size. Chernozhukov and Hansen (2006) recommend choosing a block size of $kn^{2/5}$ with k between 3 and 10. $k = 10$ gives a block size of 250. Due to the computation time only 500 replications have been drawn for the bootstrap.

6.4. Endogenous sector choice with fully interacted covariates

In Section 6.3 we have assumed that the returns to characteristics are the same in both sectors. In the exogenous case, we have 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 SIVQR of Section 3.1. Results for the MDIVQR (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 the Klein and Spady estimator (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 index estimated by Klein and Spady estimator. The results of the median regressions, given in the 3rd and 4th 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 for the other quantiles, although the p -values are 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 rejects the null hypothesis at the 0.01% level. 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 28 years of experience).

We have estimated the coefficients vectors corrected for endogeneity at 100 different quantiles uniformly distributed between 0 and 1 ($\theta = 0.005, 0.015, \dots, 0.995$). 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 5 plots the results. No confidence intervals are plotted to avoid overloading the figure. At the median, the standard errors of the estimates are 1.8%, 14%, 15% and 1.9% 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. Then, 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. While the wage premium is not significant different from zero for most of the quantiles, a joint test of the absence of a premium at the 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles rejects the null hypothesis at the 1% level.

6.5. Comparison with Abadie, Angrist and Imbens (2002) estimator

Abadie, Imbens and Angrist (2002) propose an alternative estimator based on the LATE model. They impose an independence condition (Z is independent of the errors in the outcome and selection equations) and a monotonicity condition (the direction of the effect of Z on the participation decision is the same for all individuals). They show that the marginal distributions of the potential outcome are identified for the sub-population of compliers and

suggest an ingenious estimator that can be interpreted as a re-weighted standard quantile regression estimator. The estimands of Chernozhukov and Hansen and of Abadie, Angrist and Imbens are not generally the same and the sets of assumptions are different. However, if both sets of assumptions are satisfied and if the compliers are representative of the whole population, then both estimators will converge to the same value. If these conditions are not satisfied, they will have different probability limits. Thus the comparison of the results of both estimators provides a useful robustness check.

Since the Abadie, Angrist and Imbens estimator allows only for a single instrument, we use only the most powerful instrument, the public sector status of the father. As they suggest, we implement their estimator by running weighted quantile regression. We estimate the weights by power series but, probably because of the relative small sample size, only the first order of the polynomial was found to have explanatory power. The 5th and 6th columns of Table 5 give the coefficients of the median regression for the public and the private sector respectively. The results obtained by applying the SIVQR are confirmed: the constant is higher in the public sector, returns to education and experience are higher in the private sector. Figure 6 plots the decomposition defined in Section 4 using the coefficients of 100 quantile regression estimated by the method of Abadie, Angrist and Imbens. The positive selection into the private sector appears again. The differences between the corrected and the uncorrected differential are slightly lower, maybe as a consequence of the weaker instrument which may bias the results in the direction of the results obtained by traditional quantile regression. Overall, it appears that these differences are small relatively to sampling variation and that one would not draw substantively different conclusions from either set of estimates.

6.6. *Validity of the instruments*

The crucial assumptions for the consistency of the estimators used in this paper (apart from the parametric restrictions) 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. With 5 instruments for a single endogenous variable, we can also partially test the second assumption. Since we have chosen the weighting matrix $A(\theta, \alpha)$ to be the inverse of the asymptotic covariance matrix of $\sqrt{n}(\hat{\gamma}(\theta, \alpha) - \gamma(\theta, \alpha))$, the objective function of the IVQR is asymptotically $\chi^2_{\dim(\gamma)}$ -distributed under the null-hypothesis that the exclusion restrictions are satisfied. We apply this test and can reject the null-hypothesis for none of the percentile at the 1% significance level. For instance, the value of the objective function is 6.74 at the median, which is far below standard critical values.

However, an identifying assumption cannot totally be tested and this is particularly true here because all instruments are of the same type. Therefore, we try to assess the plausibility of the exclusion restrictions. Our motivation for using these instruments is that children learn through imitation of adults living in their neighborhood, and particularly their parents. 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. Thus, the instruments we use should influence the sector choice but have no impact on the potential wages.

The exclusion restrictions would 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. This would explain why the children of a father in the public sector work over-proportionally in the public sector. As a first indication against this hypothesis, we find that the sector choice of the father is not important (and not significant) for his daughter but is strongly significant for his son. Similarly, the sector choice of their mother plays no role in the occupational choice of boys but is high and significant for girls. These results indicate that the role of the father as a model to imitate seems to be the reason for the intergenerational correlation in sector choice.

Another way of assessing the exclusion restrictions is to use information about the position of the father within the public sector. If we assume that the capacity of a public sector employee to favor his son is an increasing function of his position within the public sector hierarchy, which seems reasonable, we can test the exclusion restriction by using additional information about the position of the father. In the data we do not only know whether the father was working in the public sector or not but we also have an indication about his position within the public sector. The second column of Table 8 gives the results of the logit estimation of the sector choice equation. Instead of a single dummy for the father's public sector status we use 5 different categories related to the position of the father within the public sector. The 5 coefficients are very near and we cannot reject the hypothesis that they are equal (p -value of 0.9) but we can naturally reject the hypothesis that they are null (p -value of 0.1%). In the third column of Table 8 we give the median public sector wage gap using only one category as instrument. The level of the estimates varies somewhat but remains positive with all instruments and no "trend" can be found. Thus, nothing indicates that only sons of high-level public sector employees work more in the public sector or that they earn an higher premium if they work in the public sector.

7. Summary and conclusion

In this paper, we have examined the public / private sector wage differential allowing for the first time simultaneously for endogenous sector choice and heterogeneous public sector gaps at different points of the distribution. We have applied the instrumental quantile regression estimator of Chernozhukov and Hansen (2004b and 2006) to data from the German Socio Economic Panel which provides us with reasonable instruments. The empirical findings can be summarized as follows. 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. Thus, we find positive selection into the private sector, the sector with higher wage inequality, as predicted by the Roy (1951) model.

This paper contributes also methodologically to the literature. The estimator proposed by Chernozhukov and Hansen is convenient if the number of endogenous variables is small but becomes more and more difficult to implement when this number increases. In order to allow the public sector wage premium to vary with the covariates, we need to estimate a fully interacted model which is practically infeasible in term of computation time. Therefore, we propose two estimators that are tractable in this configuration. Their asymptotic distributions are derived and Monte-Carlo simulations show their good behavior in finite samples.

Applying the new estimators, we find that the public sector also reduces the between-group inequality by giving smaller returns to education. Thus, the government refuse to pay low wages to its less skilled employees and very high wages to its most skilled employees¹⁶. Finally, the experience-wage profile is more concave in the private sector. 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. In summary, statistical tests have shown that the sector choice is endogenous, the public sector wage premium is different at different parts of the distribution and returns to education and experience are different in the public sector.

Acknowledgements

I thank Pedro Carneiro, 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

¹⁶ This is true for observed and unobserved (for the econometrician) skills.

EALE at San Francisco, the 2005 meeting of EEA at Amsterdam and the Cost A23 conference in Paris.

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.
- 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.
- Angrist, J., Chernozhukov V. and I. Fernández-Val, 2006, Quantile Regression under Misspecification with an Application to the U.S. Wage Structure. *Econometrica*, forthcoming.
- 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.
- Carneiro, P. and S. Lee, 2005, Ability, Sorting and Wage Inequality, Cemmap working paper 16/05.
- 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, 2005, An IV Model of Quantile Treatment Effects. *Econometrica* 73, 245-262.
- Chernozhukov, V. and C. Hansen, 2006, 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* 73, 1525-1550.
- 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.
- Ichimura H., 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58, 71-120.
- Imbens G., and J. Angrist, 1994, Identification and Estimation of Local Average Treatment Effects, *Econometrica* 62, 467-476.
- Klein, R. and R. Spady, 1993, An efficient semiparametric estimator of the binary response model. *Econometrica* 61, 387-421.
- Koenker, R. and G. Bassett, 1978, Regression Quantiles. *Econometrica* 46, 33-50.
- Lee, S., 2004, Endogeneity in quantile regression models: a control function approach. CEMMAP working paper 08/04.
- Ma, L. and R. Koenker, 2006, Quantile regression methods for recursive structural equation models. *Journal of Econometrics*, forthcoming.
- Melly, B., 2005, Public-private sector wage differentials in Germany: evidence from quantile regression. *Empirical Economics* 30, 505-520.
- Melly, B., 2006, Estimation of counterfactual distributions using quantile regression. 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.
- 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, 233-257.
- 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 (2005) 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 (2006). An application can be found in Chernozhukov and Hansen (2004a).

Chernozhukov and Hansen (2004b and 2006) note that

$$\Pr(Y \leq D'\alpha(\theta) + X'\beta(\theta) | X, Z) = \theta$$

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

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

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:

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

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:

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

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

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

$\hat{A}(\theta, \alpha) = A(\theta, \alpha) + o_p(1)$ and $A(\theta, \alpha)$ is positive definite uniformly in $\alpha \in A$. It is convenient to set $A(\theta, \alpha)$ equal to the inverse of the asymptotic covariance matrix of $\sqrt{n}(\hat{\gamma}(\theta, \alpha) - \gamma(\theta, \alpha))$. In this case, more weights are given to the instruments whose effects 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 $(\hat{\alpha}(\theta), \hat{\beta}(\theta))$:

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

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

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

$J_\alpha = E[f_\varepsilon(0|X, Z, D) \Psi D']$, and $[\bar{J}_\beta', \bar{J}_\gamma']'$ is a partition of $E[f_\varepsilon(0|X, Z) \Psi \Psi' / V]^{-1}$ 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 $V^* = f_\varepsilon(0|X, Z)$ and $\Phi^* = E[D v^* | X, Z] / V^*$,

where $v^* = f_\varepsilon(0|D, X, Z)$. Then, the asymptotic variance simplifies to $\theta(1 - \theta) E[\Psi \Psi']^{-1}$

and attains the efficiency bound.

Appendix B: Asymptotic distribution of the SIVQR estimator

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)) estimators¹⁷.

2nd step: Denote by $\hat{g} = X\hat{\alpha}_X + Z\hat{\alpha}_Z$ the estimated index and by $P_S(\hat{g}) = (P_{S1}(\hat{g}), \dots, P_{SS}(\hat{g}))$ 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):

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

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

¹⁷ 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.

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

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

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

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

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

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

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

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 $\Pr(D=1|X, Z)$. 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:

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

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

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

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)¹⁸ since the true value $\alpha_0(\theta)$ and $\delta(\theta)$ solve the following moment conditions:

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

The covariance matrix of these moment conditions is given by

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

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

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

If $\beta_0(\theta)$ and $\beta_1(\theta)$ would be given and not estimated, the asymptotic variance of $(\hat{\alpha}_0(\theta), \hat{\delta}(\theta))$ 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

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

¹⁸ 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.

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 $(\hat{\alpha}_0(\theta), \hat{\delta}(\theta))$ is

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

The covariance matrix between $(\hat{\beta}_0(\theta), \hat{\beta}_1(\theta))$ and $(\hat{\alpha}_0(\theta), \hat{\delta}(\theta))$ 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 $(\Delta_{fr}, \Delta_{frx}, J_{\beta}, J_{\theta})$ are evaluated once at θ_1 and once at θ_2 .

Appendix C: Asymptotic distribution of the MDIVQR estimator

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

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

where $\Lambda_j(\tau)$ is equal to $\Lambda(\tau)$ 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 $(\hat{\alpha}_j(\tau), \hat{\delta}_j(\tau))$ is independent of $(\hat{\alpha}_{j'}(\tau), \hat{\delta}_{j'}(\tau))$ for $j \neq j'$.

In order to obtain a condensed presentation of the results, we can now use the minimum distance framework to estimate the global parameters. Recall that we assume that the conditional quantiles of Y given X are linear in each sector. That is,

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

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

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

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

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

Then

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

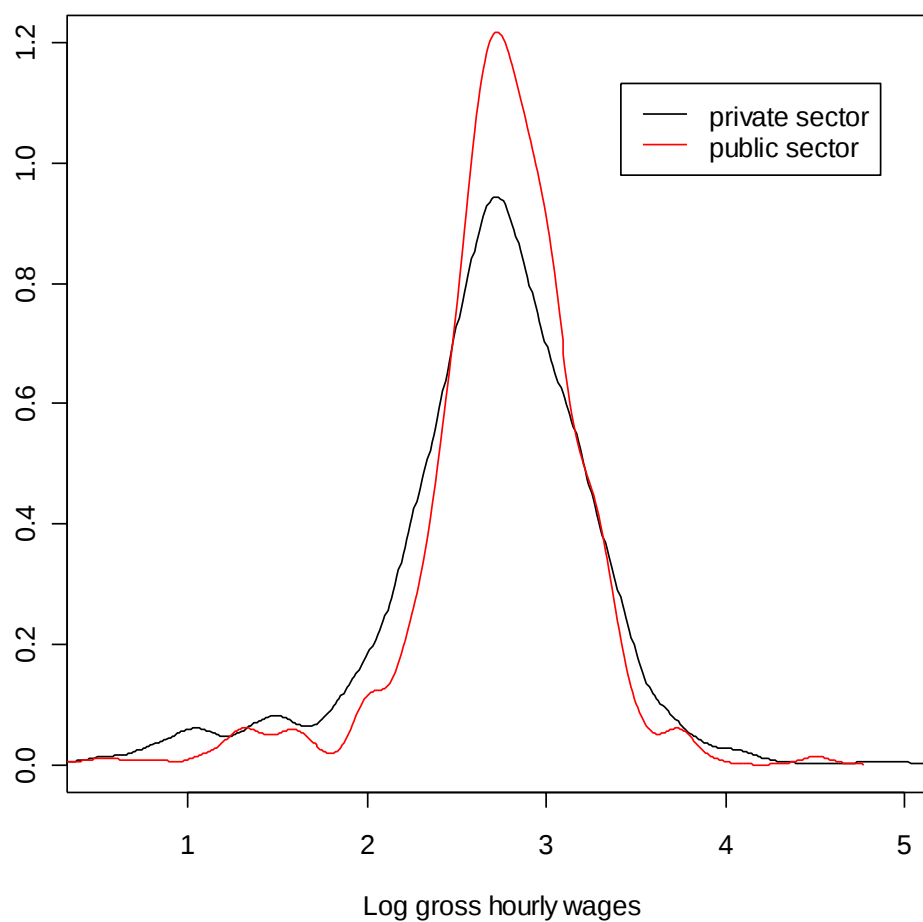
for $s = 0, 1$, where $\Omega_s(\tau)$ is a J diagonal matrix with the j^{th} diagonal element equal to the variance of $\hat{\alpha}_j(\tau)$ and $\Omega_1(\tau)$ is a J diagonal matrix with the j^{th} diagonal element equal to the variance $\hat{\alpha}_j(\tau) + \hat{\delta}_j(\tau)$. An efficient minimum distance estimator is obtained by setting $W_s(\tau)$ equal to a consistent estimator of Ω_s^{-1} . Note that the efficiently weighted minimum

distance estimator is asymptotically equivalent to the estimator of Chernozhukov and Hansen with optimal instruments and weights if the model is correctly specified. If the model is misspecified, they will generally converge to different value since they weight differently the different points in the support of X .¹⁹

¹⁹ See Angrist, Chernozhukov and Fernández-Val (2006) for a discussion of misspecified quantile regression.

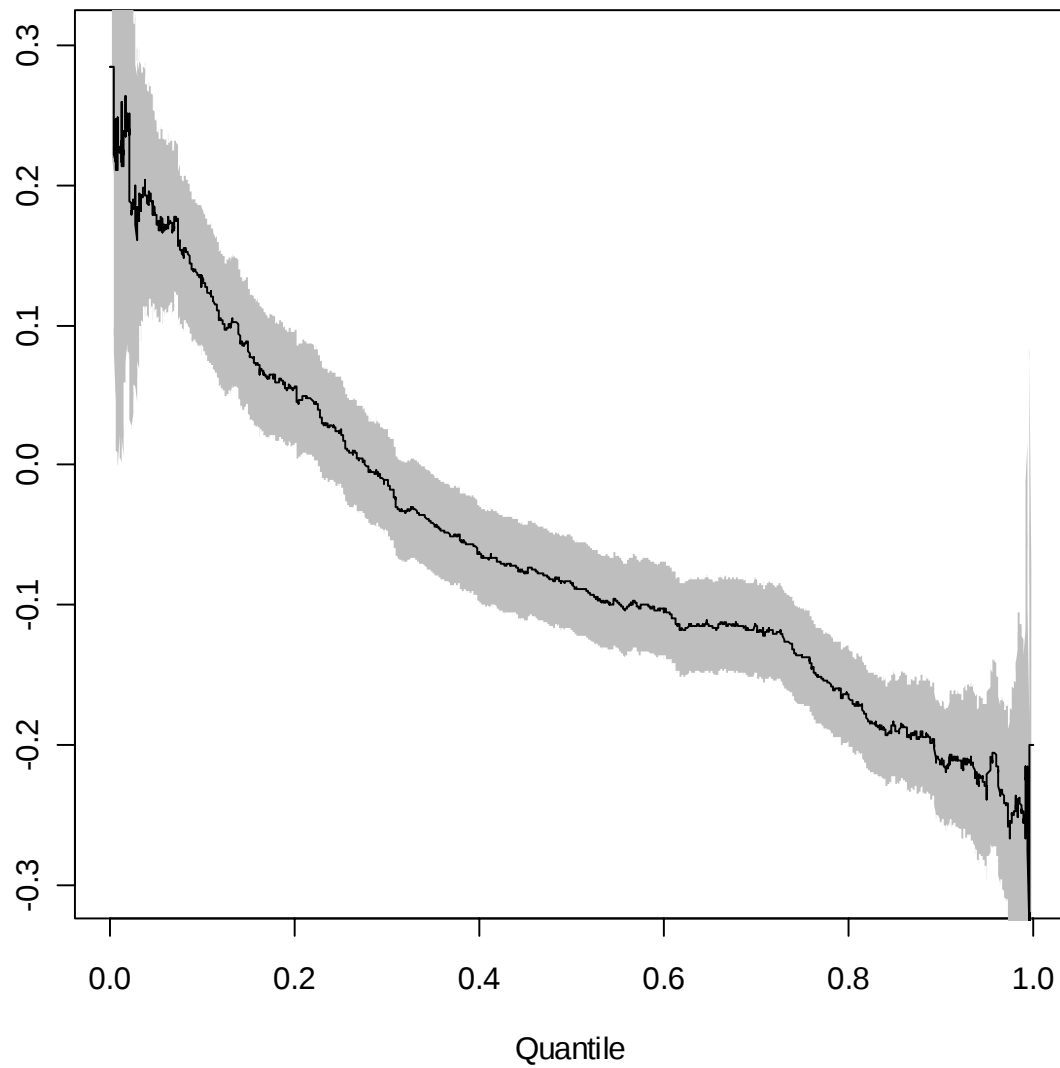
Figures

Figure 1: Kernel density estimates of the wage distributions



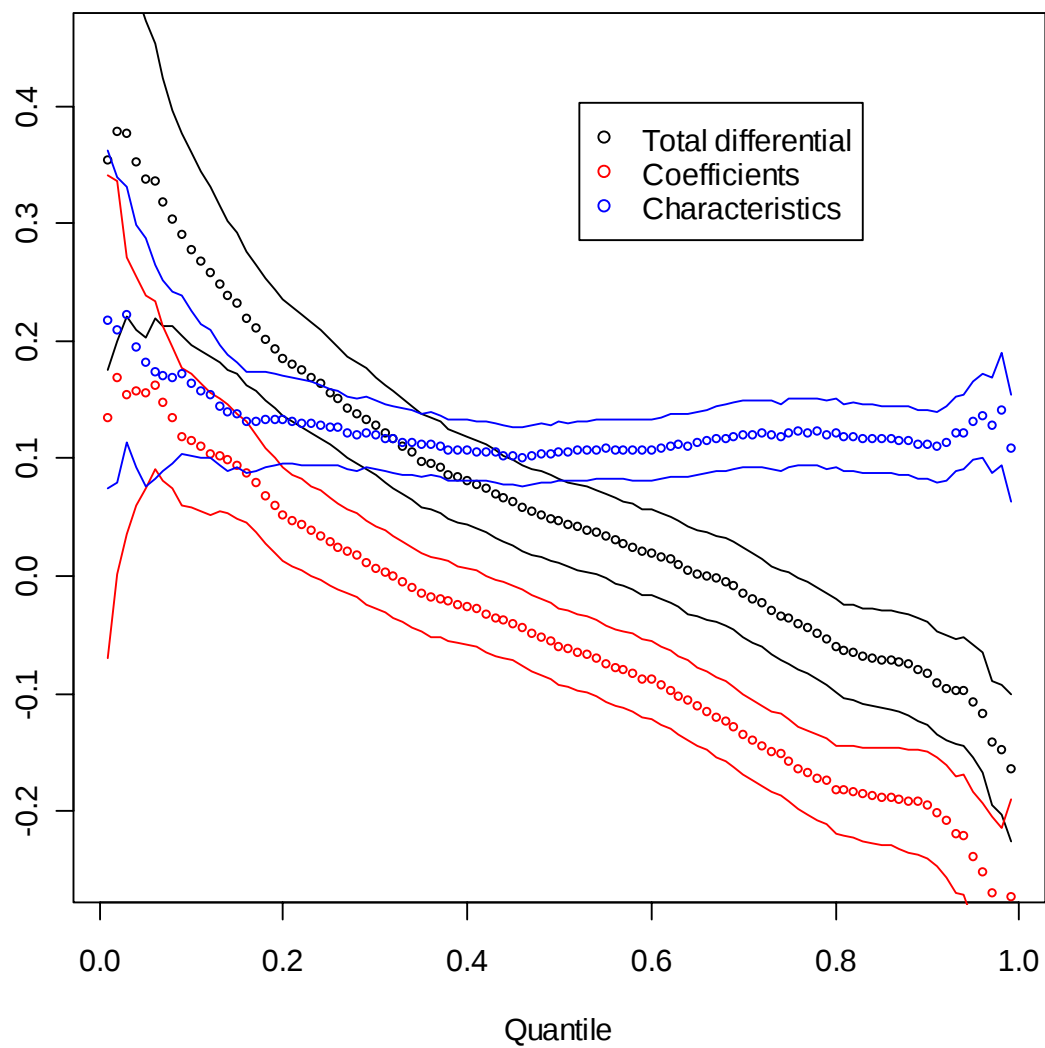
Note: Density functions estimated using an Epanechnikov kernel estimator and bandwidths chosen according Silverman's rule of thumb.

Figure 2: Public sector wage “premium” at different quantiles



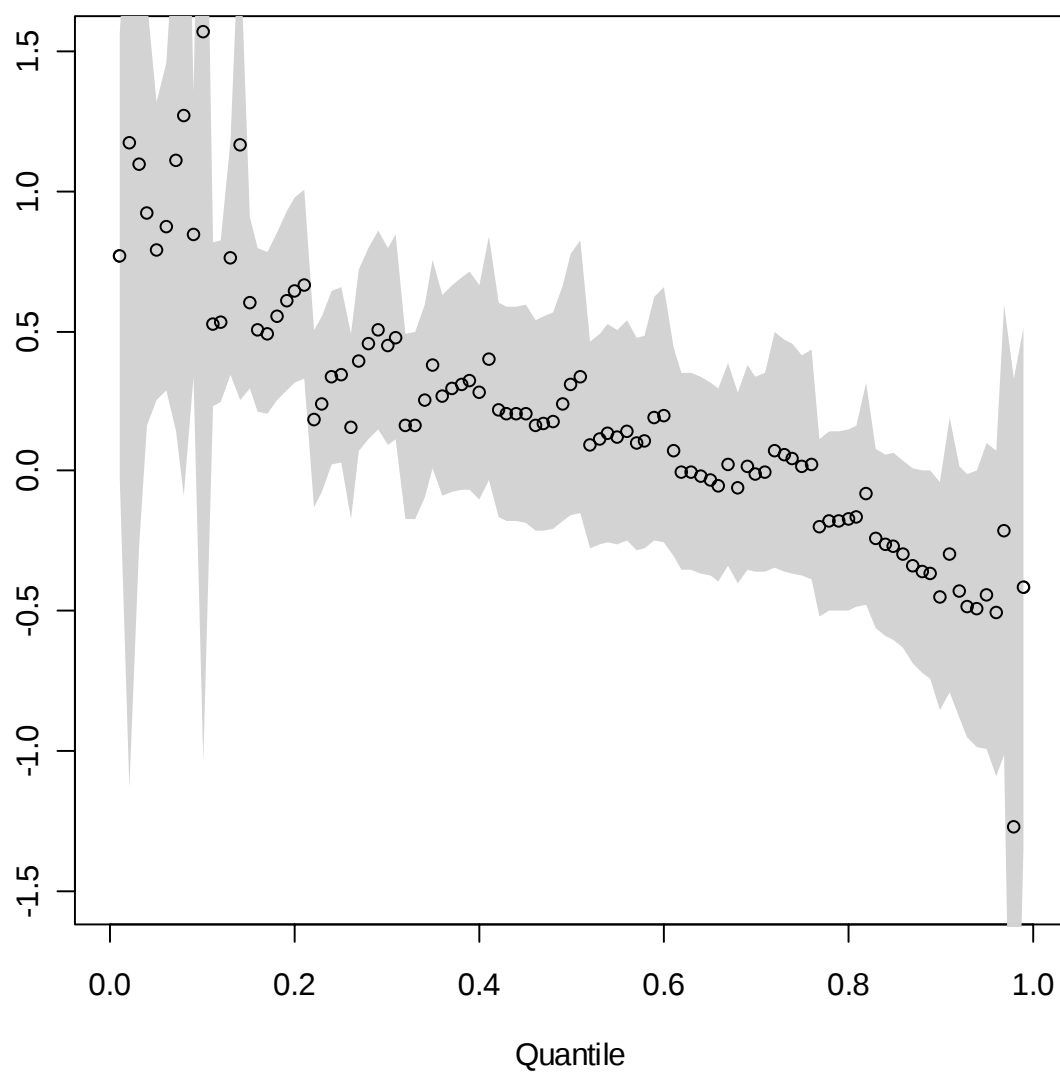
Coefficient on the public sector dummy variable estimated by traditional quantile regression with a 95% confidence interval.

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



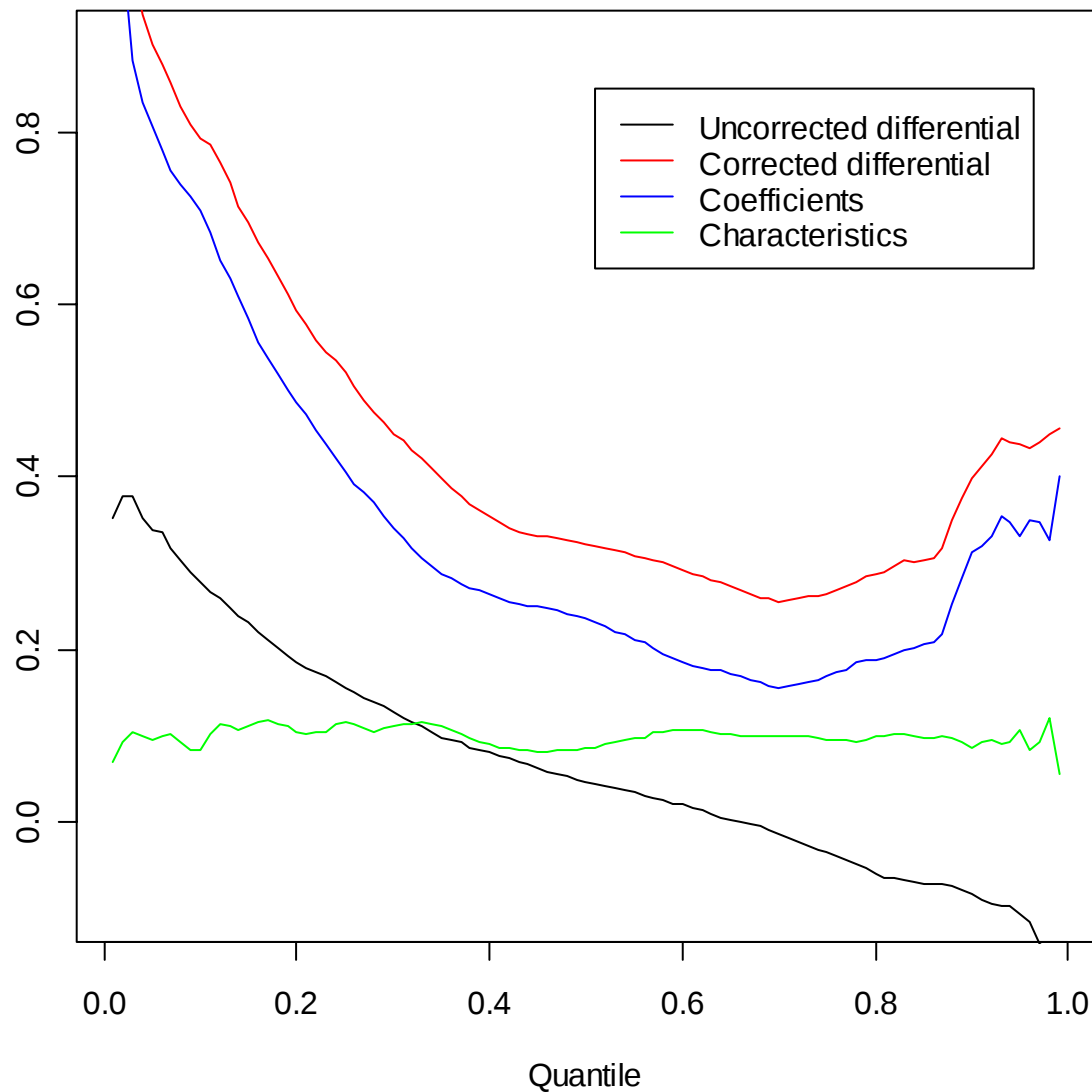
Decomposition defined in Section 4 with 95% confidence intervals delimited by the lines.

Figure 4: Public sector wage premium using instrumental quantile regression



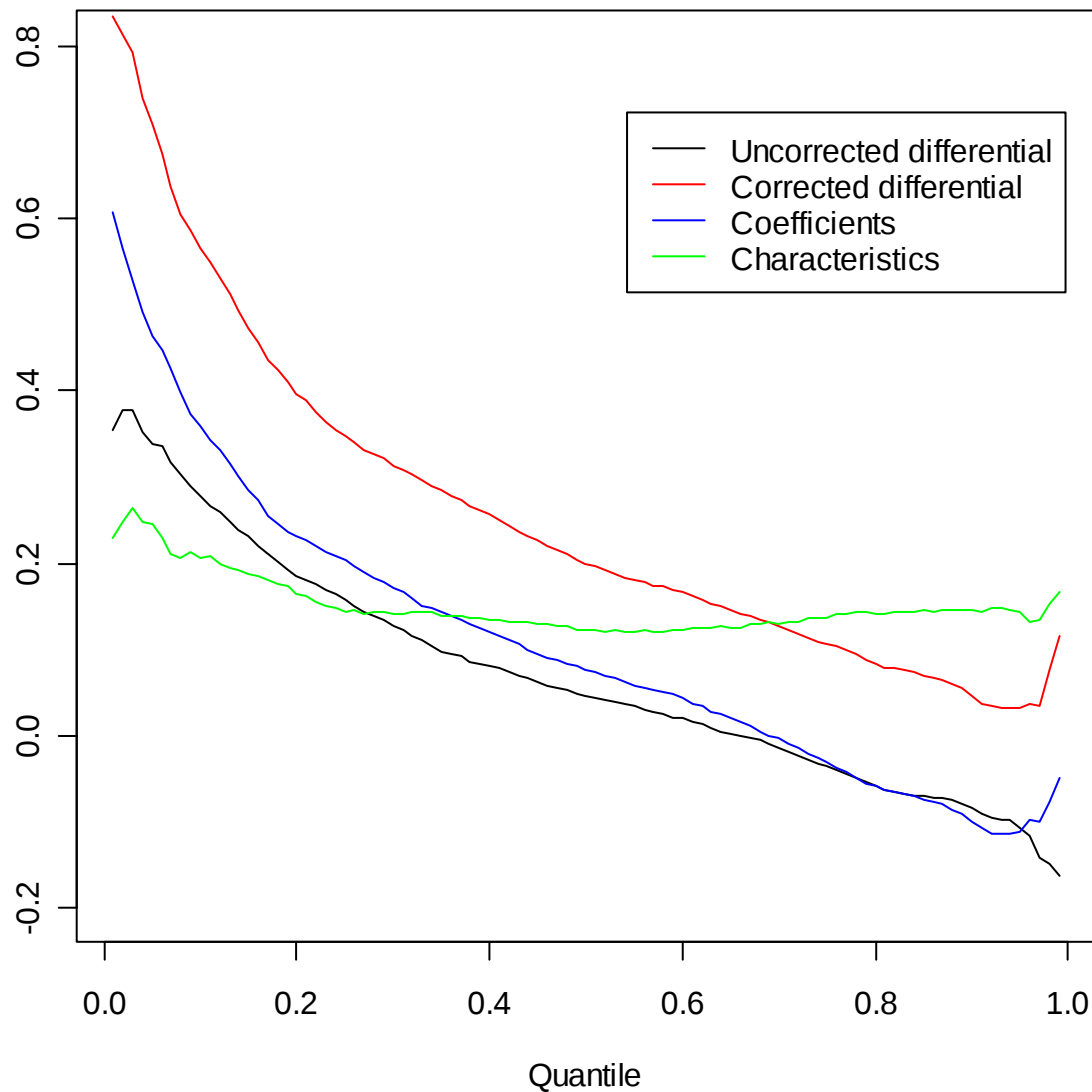
Instrumental variable quantile regression. Coefficient on the public sector dummy variable with a 95% confidence interval.

Figure 5: Decomposition of public private sector wage differential correcting for endogeneity



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 have prevailed if the sector status were exogenous. All coefficients used to estimate the counterfactual distribution were estimated by the SIVQR estimator.

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



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 status were exogenous. All coefficients used to estimate the counterfactual distribution were estimated by the AAI estimator.

Tables

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	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE
100 observations												
QR	-0.81	0.3	0.74	0.33	0.3	0.2	1.19	0.3	1.52			
IVQR	0.03	0.46	0.21	-0.04	0.38	0.15	-0.02	0.54	0.29			
IVQR	0.06	0.78	0.62	-0.09	0.87	0.77	-0.05	0.93	0.87	0.04	1.29	1.66
SIVQR	-0.01	0.51	0.26	-0.02	0.54	0.3	0.03	0.64	0.41	0	0.64	0.41
MDIVQR	0.04	0.63	0.39	-0.05	0.72	0.52	-0.02	0.78	0.61	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			
IVQR	0.01	0.21	0.04	-0.01	0.17	0.03	0	0.25	0.06			
IVQR	0.02	0.3	0.09	-0.02	0.35	0.12	-0.01	0.38	0.15	0	0.52	0.27
SIVQR	0	0.23	0.05	-0.01	0.25	0.06	0.01	0.29	0.09	0	0.29	0.09
MDIVQR	0.02	0.3	0.09	-0.02	0.34	0.12	-0.01	0.37	0.14	0.01	0.51	0.26
1600 observations												
QR	-0.81	0.07	0.65	0.33	0.07	0.12	1.18	0.07	1.4			
IVQR	0	0.1	0.01	0	0.08	0.01	0	0.12	0.02			
IVQR	0	0.14	0.02	0	0.17	0.03	0	0.18	0.03	0	0.25	0.06
SIVQR	0	0.11	0.01	0	0.12	0.02	0	0.15	0.02	0	0.15	0.02
MDIVQR	0	0.14	0.02	0	0.16	0.03	0	0.18	0.03	0	0.25	0.06

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

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	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE
100 observations												
QR	-0.8	0.38	0.78	0.32	0.47	0.32	1.19	0.44	1.6	0	0.6	0.36
IVQR	0.42	0.52	0.45	-0.61	0.4	0.53	-0.55	0.57	0.62			
IVQR	0.06	0.77	0.6	-0.09	0.87	0.77	-0.04	0.92	0.85	0.05	1.28	1.65
SIVQR	-0.01	0.51	0.26	-0.02	0.54	0.3	0.03	0.64	0.41	0	0.64	0.41
MDIVQR	0.04	0.62	0.39	-0.07	0.72	0.52	-0.02	0.78	0.61	0.08	1.03	1.07
400 observations												
QR	-0.82	0.19	0.71	0.36	0.23	0.18	1.2	0.22	1.5	-0.05	0.3	0.09
IVQR	0.39	0.25	0.22	-0.58	0.19	0.37	-0.53	0.27	0.36			
IVQR	0.01	0.3	0.09	-0.01	0.34	0.12	-0.01	0.38	0.14	0	0.52	0.27
SIVQR	0	0.23	0.05	-0.01	0.25	0.06	0.01	0.29	0.09	0	0.29	0.09
MDIVQR	0.02	0.3	0.09	-0.02	0.34	0.12	-0.01	0.37	0.14	0.01	0.51	0.26
1600 observations												
QR	-0.82	0.09	0.69	0.36	0.12	0.14	1.21	0.11	1.47	-0.06	0.15	0.02
IVQR	0.38	0.12	0.16	-0.57	0.09	0.33	-0.53	0.13	0.3			
IVQR	0	0.14	0.02	0	0.17	0.03	0	0.18	0.03	0	0.25	0.06
SIVQR	0	0.11	0.01	0	0.12	0.02	0	0.15	0.02	0	0.15	0.02
MDIVQR	0	0.14	0.02	0	0.16	0.03	0	0.18	0.03	0	0.25	0.06

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

Table 3: Definition of the variables

<i>Variable</i>	<i>Description</i>
Lnghwage	The natural logarithm of gross hourly earnings from employment. Gross hourly wage are derived by dividing gross monthly earnings by monthly actual hours worked.
Expr	Number of years of potential work experience the individual has accumulated. It is measured by $\min(\text{age} - \text{schooling} - 6, \text{age} - 18)$.
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..

Table 4: Descriptive statistics, means

<i>Variable</i>	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

Table 5: Median regression using different estimators

	RQ	IVQR	SIVQR public	SIVQR private	AAI public	AAI private
Constant	1.614** (0.056)	1.522** (0.064)	1.961** (0.244)	1.601** (0.066)	1.732** (0.138)	1.170** (0.378)
Expr	0.062** (0.004)	0.063** (0.005)	0.051** (0.009)	0.060** (0.006)	0.052** (0.007)	0.0811** (0.017)
Expr^2	-1e-3** (9e-5)	-0.001** (1e-4)	-7e-4** (2e-4)	-1e-3** (1e-4)	-7e-4** (2e-4)	-1e-3** (3e-4)
Ed level 2	0.198** (0.041)	0.26** (0.047)	0.085 (0.094)	0.251** (0.050)	0.089 (0.087)	0.268** (0.087)
Ed level 3	0.317** (0.042)	0.32** (0.049)	0.187* (0.094)	0.307** (0.053)	0.196* (0.087)	0.427** (0.112)
Ed level 4	0.34** (0.053)	0.322** (0.07)	0.256* (0.114)	0.254* (0.077)	0.295** (0.095)	0.413** (0.093)
Ed level 5	0.587** (0.045)	0.587** (0.055)	0.406** (0.104)	0.594** (0.064)	0.435** (0.087)	0.717** (0.104)
Ed level 6	0.709** (0.045)	0.659** (0.069)	0.481** (0.112)	0.684** (0.079)	0.525** (0.084)	0.848** (0.112)
Psect	-0.084** (0.02)	0.31 (0.237)				

Column 1: exogenous quantile regression, column 2: instrumental variable quantile regression, column 4: SIVQR in the public sector, column 5: SIVQR in the private sector, column 6: AAI in the public sector, column 7: AAI in the private sector. *: significant at the 5% level, **: significant at the 1% level. Analytical heteroscedasticity consistent standard errors are given in parenthesis.

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

	Logit		Klein and Spady	
	Coefficient	Std. error	Coefficient	Std. error
Constant	-1.6247***	0.2195		
Expr	0.0240**	0.0120	0.0270**	0.0143
Expr^2	-0.0001	0.0002	-0.0001	0.0003
Ed level 2	-0.0218	0.1185	-0.1711	0.1748
Ed level 3	0.3014***	0.0806	0.2229***	0.1316
Ed level 4	0.4449***	0.0922	0.3964***	0.1118
Ed level 5	0.3469***	0.0876	0.3352***	0.1230
Ed level 6	0.5863***	0.0611	0.6126***	0.1005
Fcivil	0.4182***	0.1038	0.4767***	0.1246
Mnwork	0.1337**	0.0646	0.1705**	0.0791
Fblue	0.1079	0.0754	0.0120	0.1015
Fself	0.1307	0.0927	0.0514	0.0966
Fwhite	0.1569*	0.0853	0.1287	0.1017

Standard errors are obtained by bootstrapping the results 100 times for the Klein and Spady estimator and 1000 times for the logit estimator. *: significant at the 10%, **:significant at the 5%, ***: significant at the 1%.

Table 7: *P*-values on the instrumental quantile regression process

Null hypothesis	Subsample size		
	250	2000	3125
No effect: $\alpha(\cdot) = 0$	<0.1%	3%	4.6%
Constant effect: $\alpha(\cdot) = \alpha$	<0.1%	2.7%	2.8%
Exogeneity: $\alpha(\cdot) = \alpha_{QR}(\cdot)$	<0.1%	3.1%	3.8%
Dominance: $\alpha(\cdot) \geq 0$	36.3%	57.1%	56.4%
Dominance: $\alpha(\cdot) \leq 0$	<0.1%	0.1%	2.2%

The statistics and critical values are computed using the method of Chernozhukov and Hansen (2006) using the Smirnov-Cramer-Von Misses statistic. 1000 replications with subsample sizes of 250 or 2000 observations and 500 replications with the bootstrap were constructed.

Table 8: Results disaggregated by the position of the father within the public sector

Hierarchical level of the father within the public sector	Coefficient in the logit estimation of the sector choice equation	Median public sector wage premium using only one category as instrument
Low-level	0.640 (0.430)	0.078 (0.207)
Middle-level	0.734*** (0.226)	0.306 (0.274)
High-level	0.864*** (0.261)	0.118 (0.204)
Executive	0.532* (0.284)	0.082 (0.218)
No information ²⁰	0.540 (0.487)	0.115 (0.203)

*, significant at the 5% level, **, significant at the 1% level. Analytical heteroscedasticity consistent standard errors are given in parenthesis. The second column gives the results gives the coefficients on the public sector status of the father obtained by the logit estimation of the sector choice equation. The other covariates are the same as those used in Table 6. The third column give the median public sector wage premium estimated by IVQR with only one hierarchical level as instruments.

²⁰ These are public sector employees without information about the level in the hierarchy.