Quantile-based inference of parametric transformations between two distributions*

Laurent Gobillon (INED)† Sébastien Roux (CREST-INSEE)‡

February 28, 2008
Preliminary Version

Abstract

In this paper, we compare the distributions of a continuous outcome between two groups. We focus on specifications such that the two distributions are related by some parametrized transformations of values and ranks. These specifications can be derived from a specific class of theoretical models. We propose a quantile method to estimate the parameters and derive their asymptotic distribution. We also propose a test of the specification. Monte-Carlo simulations show that the estimators of the parameters perform very well when the number of observations in each group is at least a few thousands. We finally apply our method to the wage differential between skilled males and females in a selection of sectors. We are able to quantify the respective effects of the glass ceiling and a uniform discrimination which may lower the wages of females.

Keywords: quantiles, distributions, model estimation
JEL Classification: C12, C24, C52

*We are grateful to Stéphane Grégoir, Marc Gurgand, Edwin Leuven and Dominique Meurs for useful comments and discussions, as well as the participants of the microeconometrics seminar at CREST.

†Institut National d'Etudes Démographiques (INED), 133 Boulevard Davout, 75980 Paris Cedex 20, France. Email: laurent.gobillon@ined.fr.

‡Centre de Recherche en Economie et Statistique (CREST), 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. Email: sebastien.roux@ensae.fr.

1 Introduction

A growing literature is interested in comparing the distributions of a continuous outcome between two groups. The most important econometric contributions are semi or non parametric in spirit and try to assess the effect of a policy on the whole distribution of outcomes. They assess the impact of a policy at a given quantile comparing the distributions of the treated and untreated groups. This type of approach was initiated by Doksum (1974) who estimated the effect of a treatment at a given rank of the distribution simply as the difference between the quantiles of the treated and untreated groups computed at this rank. Papers now introduce explanatory variables in models and propose IV methods (Abadie, Angrist and Imbens, 2002; Chernozhukov and Hansen, 2005), propensity score methods (Firpo, 2007), or difference-in-difference improvements (Athey and Imbens, 2006). This literature typically tries to estimate the effect of the treatment under minimum parametric conditions.

In this paper, we are rather interested in estimating and testing a theoretical model under some restrictive parametrizations. We will focus on a given class of models which result in the distributions of the two groups to be linked by some transformations of values and ranks. If we denote λ_j the quantile function of group j, these models yield a relationship that writes: $\lambda_2 \left[r_{\gamma_0}(u)\right] = z_{\beta_0}\left[\lambda_1(u)\right]$ for every quantile u, where β_0 and γ_0 are some parameters, z_{β_0} is the transformation of values, and r_{γ_0} is the transformation of ranks. We also asume that the value and rank transformations are monotonic. Several types of micro-economic models can give rise to this specification. In particular, this specification is obtained when the continuous outcome is observed only for a selected group of agents and the probability to be selected is a monotonic function of the ranks of the agent in the distribution of their group.

We develop a method to estimate the parameters of the model and derive the asymptotic distribution of the estimators. The estimation method can be decomposed into two steps. We first estimate the quantiles of the two group distributions non parametrically. We then minimize the distance between the quantiles of the two groups after the value and rank transformations have been applied. Note that even if our method uses quantiles, our work is not directly related to the literature on quantile regressions (see Köenker, 2005 for an overview). Indeed, we minimize the distance between transformed quantile functions rather than the penalized sum of absolute deviations at the individuel level. We then propose a test of the specification of the model which allows to assess whether the mechanisms included in the model are compatible with the data. Some Monte-Carlo simulations show that our method works very well provided that there are enough observations in the two groups, that is to say at least a few thousands.

Our method can be applied in many settings. In particular, many theoretical papers develop models with agents who are heterogeneous in a given dimension. *Ex-ante*, the group distributions of agents are assumed to be the same. Distributions are then transformed by some economic mechanisms. *Ex-post*, the group distributions usually differ but may be related by some value and rank transformations. This type of model can by found for instance in international trade when looking at the effect of competition on the local distributions of firm productivities. Competition is usually fiercer in larger markets, which makes a larger share of firms at the bottom of the local productivity distribution to disappear (Melitz, 2003; Melitz and Ottaviano, 2007). This left truncation corresponds

to a rank transformation. Other papers put forward the existence of more interactions in larger markets which may benefit to all firms (Duranton and Puga, 2004). These interactions can be modelled as a translation of the productivity distribution in a large market to the right compared to the distribution in a small market. This translation corresponds to a value transformation. Our method allows to assess whether competition or interactions prevail by comparing the productivity distributions of firms in small and large markets (Combes and al., 2008).
An originality of our work is that we use the whole group distributions to test a theory. This is not what is usually done in the empirical literature which rather uses fragmented information on the group distributions. For instance, papers testing the glass ceiling effect often run separate quantile regressions for wages at different quantile values.[1] Each regression includes a dummy for being a female. The papers then consider that there is a glass ceiling if the gender gap is larger at the top of the distribution than at the bottom (Albrecht, Björklund and Vroman, 2003). In an application, we parametrize the effects of the glass ceiling and a uniform discrimination against females. We estimate the parameters of these two discriminations comparing on the whole wage distributions of males and females. This is done in one single minimization program using all the quantiles of the two distributions simultaneously. We find that both types of discrimination have a significant effect on the wages of females.

In the first section, we expose our setting where two group distributions are related to each other through some parametrized rank and value transformations. Section two proposes some examples of transformations to which our approach can be applied. Section three presents our estimation method and derive the asymptotic distribution of the estimators. It also provides a test of the model. Section four explains how our method can be implemented in an easy way. Results of Monte-Carlo simulations are given in section five. We then propose an application of the method to the difference in the wage distribution between males and females in section 6. Finally, section 7 concludes.

2 The setting

We consider two groups of heterogeneous agents who differ in a given dimension measured by a variable y which takes its values in a bounded interval included in $\mathbb{R}$. The draws of y for the agents are independent. We denote F_j the cumulative of the outcome distribution in group j. We assume that the distribution in one of the two groups (say 2) can be deduced from the distribution in the other group (say 1) through some parametrized value and rank transformations. More precisely, while a monotonic function $z_{\beta_0,2}(y)$ transforms the values of the outcome variable y, another monotonic function $r_{\gamma_0,2}(u)$ changes the ranks u of the transformed outcome. Here, β_0 and γ_0 are some parameters of interest that we want to estimate. The functions are indexed by the subscript 2 as we write the distribution in the second group as a transformation of the distribution in the first group (we will show that under our assumptions, we may swap the two groups). We are going to look for the parameters of interest (β_0, γ_0) in a set of admissible parameters (β, γ) which we denote Φ. We make the following assumptions:

[1] Other examples include Juhn, Murphy and Pierce (1993) who study the variation across time in the wage distribution of males in the US, and Syverson (2004).

A1: For $(\beta, \gamma) \in \Phi$, $r_{\gamma,2}$ and $z_{\beta,2}$ are strictly increasing and three times differentiable. The first, second and third-order derivatives of $r_{\gamma,2}$ and $z_{\beta,2}$ are continuous.

In particular, *A1* ensures that the two functions $r_{\gamma,2}$ and $z_{\beta,2}$ are invertible, which will be used extensively in the paper. We denote $z_{\beta,1} = z_{\beta,2}^{-1}$ and $r_{\gamma,1} = r_{\gamma,2}^{-1}$. In fact, *A1* also assume that the two functions are increasing, but the decreasing case can be treated similarly. More formally, the link between the two distributions is written as:

$$F_2(z_{\beta_0,2}(y)) = r_{\gamma_0,2}(F_1(y)) \text{ for } F_1(y) \in r_{\gamma_0,1}([0,1]) \cap [0,1] \tag{1}$$

Here, $z_{\beta_0,2}$ changes the y-values and $r_{\gamma_0,2}$ transforms some ranks of the first distribution into some ranks of the second distribution. $r_{\gamma_0,1}([0,1]) \cap [0,1]$ is the set of ranks in the first distribution for which (1) holds. $[0,1] \cap r_{\gamma_0,2}([0,1])$ is the set of ranks in the transformed second distribution for which (1) holds. Note that the set of ranks in distribution j which participate to (1) rewrite as $\left[\underline{u}_{\gamma_0,j}, \overline{u}_{\gamma_0,j}\right]$ where $\underline{u}_{\gamma,j} = \max(0, r_{\gamma,j}(0))$ and $\overline{u}_{\gamma,j} = \min(1, r_{\gamma,j}(1))$. In particular, we can distinguish the three following cases:

- $r_{\gamma_0,1}(0) > 0$. We have $\underline{u}_{\gamma_0,1} > 0$ and the lowest ranks of the first distribution are not used. We also have: $r_{\gamma_0,2}\left(\underline{u}_{\gamma_0,1}\right) = 0$ which means that the lowest ranks of the transformed second distribution are used.
- $r_{\gamma_0,1}(0) = 0$. We have $\underline{u}_{\gamma_0,1} = 0$ and $r_{\gamma_0,2}\left(\underline{u}_{\gamma_0,1}\right) = 0$ which means that the lowest ranks of both the first and the transformed second distributions are used.
- $r_{\gamma_0,1}(0) < 0$. We have $\underline{u}_{\gamma_0,1} = 0$ and the lowest ranks of the first distribution are used. We also have: $r_{\gamma_0,2}\left(\underline{u}_{\gamma_0,1}\right) = r_{\gamma_0,2}(0) > 0$ which means that the lowest ranks of the transformed second distribution are not used.

The same line of arguments can be applied at the top of the distributions looking at $r_{\gamma_0,1}(1) \lesseqgtr 1$. Note that $r_{\gamma_0,1}$ changes the admissible ranks non uniformely and can account for the disappearance of some observations depending on their y-value. This means that it can account for selection. We want to rewrite equation (1) with quantiles and we will need quantities to be differentiable in the estimation section. Hence, we make the following assumption:

A2: for $j \in \{1,2\}$, F_j is strictly increasing and three time differentiable. The first, second and third-order derivatives of F_j are continuous.

Assumption A2 makes F_j invertible. Let $u = F_1(y)$ and $\lambda_j = F_j^{-1}$, equation (1) can be rewritten as:

$$z_{\beta_0,2}(\lambda_1(u)) = \lambda_2\left(r_{\gamma_0,2}(u)\right), \text{ for } u \in \left[\underline{u}_{\gamma_0,1}, \overline{u}_{\gamma_0,1}\right] \tag{2}$$

The set of equations (2) constitute an infinite set of equalities between transformed quantiles from which it is possible to identify the parameters β_0 and γ_0. We suppose that the solution is unique:

A3: There is a unique couple (β_0,γ_0) which verifies (2).

When specifying the relationships between the distributions in the two groups, we treated these groups asymetrically. In fact, it is possible to swap their role. We can apply the function $z_{\beta_0,1}$ to both sides of the equalities (2) compute at the ranks $r_{\gamma_0,1}(u)$. We obtain:

$$\lambda_1\left(r_{\gamma_0,1}(u)\right) = z_{\beta_0,1}\left(\lambda_2(u)\right), \text{ for } u \in \left[\underline{u}_{\gamma_0,2}, \overline{u}_{\gamma_0,2}\right] \tag{3}$$

These equations are the symetric of (2). We will use both sets of equations (2) and (3) to estimate the parameters.

3 Interpretations

3.1 Deriving the specification from a theoretical model

We now show that the equalities (2) can be derived from a class of theoretical models with two groups. This class is defined such that before any economic mechanism takes place, the distribution of potential outcomes in the two groups are the same. We denote $\tilde{F}$ the cumulative of this common distribution which we call the *baseline distribution*. The theoretical model transforms the cumulative $\tilde{F}$ through some economic mechanisms into a cumulative F_1 (resp. F_2) for the first (resp. second) group. For instance, the baseline distribution can be the distribution of productivities of firms before they enter a small or a large market. The less productive entrants disappear because of competition which is fierce on the large market and less intense on the small one (see Melitz, 2003). The distribution of firm productivities on the two markets after competition (which are the ones observed) are not the same as more firms disappear form the large market. They also differ from the baseline distribution. We assume that the distribution of each group j can be derived from the baseline distribution through some value and rank tranformations denoted respectively $z_{\beta_j}(y)$ and $r_{\gamma_j}(y)$. We make the assumption that:

A1b: r_{γ_j} and z_{β_j} are strictly increasing and three times differentiable. The first, second and third-order derivatives of r_{γ_j} and z_{β_j} are continuous.
A2b: $\widetilde{F}$ is strictly increasing and three time differentiable. The first, second and third-order derivatives of $\widetilde{F}$ are continuous.

Here, *A1b and A2b* ensure that the functions $\widetilde{F}$, r_{γ_j} and z_{β_j} are invertible. The link between the group distributions and the baseline distribution is given by:

$$F_j(z_{\beta_j}(y)) = r_{\gamma_j}\left(\widetilde{F}(y)\right) \text{ for } F_1(y) \in \left[r_{\gamma_j}^{-1}(0), r_{\gamma_j}^{-1}(1)\right]$$

This relationship is simpler than (1). Indeed, the group distributions may be a truncated version of the baseline distribution, but the reverse cannot be true. We denote $\widetilde{\lambda} = \widetilde{F}^{-1}$. The transformed quantile functions verify:

$$z_{\beta_j}\left(\widetilde{\lambda}(u)\right) = \lambda_j\left(r_{\gamma_j}(u)\right), \text{ for } u \in \left[r_{\gamma_j}^{-1}(0), r_{\gamma_j}^{-1}(1)\right] \tag{4}$$

Note that this specification is similar to the one introduced in the previous section which lead to (2). Applying the function $z_{\beta_j}^{-1}$ on both sides of (4), we get:

$$\widetilde{\lambda}(u) = z_{\beta_j}^{-1}\left[\lambda_j\left(r_{\gamma_j}(u)\right)\right], \text{ for } u \in \left[r_{\gamma_j}^{-1}(0), r_{\gamma_j}^{-1}(1)\right] \tag{5}$$

Denote $\Theta = \left\{ (\gamma_1, \gamma_2) \left| [r_{\gamma_1}^{-1}(0), r_{\gamma_1}^{-1}(1)] \cap [r_{\gamma_2}^{-1}(0), r_{\gamma_2}^{-1}(1)] \neq \oslash \right. \right\}$. The data do not bring any information to identify the parameters of the model for values of (γ_1, γ_2) which do not belong to Θ. Indeed, for these values, equation (5) is verified at best only for $j = 1$ or $j = 2$ for any quantile y and the function $\tilde{\lambda}(u)$ is not observed. Hence from now on, we will focus only on parameters γ_1 and γ_2 in the set of values Θ. Using (5), we can then write that:

$$z_{\beta_1}^{-1}\left(\lambda_1\left(r_{\gamma_1}(u)\right)\right) = z_{\beta_2}^{-1}\left(\lambda_2\left(r_{\gamma_2}(u)\right)\right), \text{ for } u \in \left[\max\left(r_{\gamma_1}^{-1}(0), r_{\gamma_2}^{-1}(0)\right), \min\left(r_{\gamma_1}^{-1}(1), r_{\gamma_2}^{-1}(1)\right)\right]$$

These equations rewrite:

$$\left(z_{\beta_2} \circ z_{\beta_1}^{-1}\right)(\lambda_1(u)) = \lambda_2\left(\left(r_{\gamma_2} \circ r_{\gamma_1}^{-1}\right)(u)\right), \text{ for } u \in \left[\max\left(0, r_{\gamma_1} \circ r_{\gamma_2}^{-1}(0)\right), \min\left(1, r_{\gamma_1} \circ r_{\gamma_2}^{-1}(1)\right)\right]$$

Denote $z_{\beta_0,2} = z_{\beta_2} \circ z_{\beta_1}^{-1}$, $r_{\gamma_0,2} = r_{\gamma_2} \circ r_{\gamma_1}^{-1}$, $z_{\beta_0,1} = z_{\beta_0,2}^{-1}$ and $r_{\gamma_0,1} = r_{\gamma_0,2}^{-1}$, where β_0 (resp. γ_0) are some combinations of the parameters β_1 and β_2 (resp. γ_1 and γ_2) which are identified. We are back to the specification in the first section. This means that the specification yielding (1) corresponds to a theoretical model where only some combinations of parameters are identified. For instance, if the rank transformation is the identity and the value transformation corresponds to a transaltion ($z_{\beta j}(y) = y - \beta_j$), it is possible to identify only the difference in translation between the two groups ($\beta_2 - \beta_1$). Note that assumptions *A1b* and *A2b* are sufficient for *A1* and *A2* to be verified. This means that assumptions on the underlying theoretical model may be used to grant assumptions needed for the reduced form given by (2). In the next subsection, we give some more examples of specifications for the functions $z_{\beta j}$ and r_{γ_j} which lead to equalities of the form (2).

3.2 Examples

In this subsection, we characterize the functions z_{β_j} and r_{γ_j} for two transformations which are quite common in the literature: a linear transformation of the $y-$values and selection. We show how to deduce the functions z_{β_0} and r_{γ_0} of the reduced form (2). Of course, the resduced form could be used directly, without referring to any underlying theoretical model.

1. The linear transformation is charaterized by $z_{(a_j,b_j)}(y) = a_j y + b_j$ for $j \in \{1, 2\}$. In that case, the relative linear transformation between groups 1 and 2 writes:

$$\begin{aligned} z_{\beta_0,2}(y) &= \left(z_{(a_2,b_2)} \circ z_{(a_1,b_1)}^{-1}\right)(y) \\ &= z_{(a_2,b_2)}\left(\frac{x - b_1}{a_1}\right) = \frac{a_2}{a_1}y + b_2 - \frac{a_2}{a_1}b_1 \\ &= a_0 y + b_0 \end{aligned}$$

   where $a_0 = \frac{a_2}{a_1}$, $b_0 = b_2 - \frac{a_2}{a_1}b_1$, and $\beta_0 = (a_0, b_0)'$.

2. We now consider the case where the distributions of groups 1 and 2 are some transformations of the baseline distribution through a selection process. We assume that for each group j, the observations are selected with respect to their rank u in the baseline distribution according to a sampling function $p_{\gamma_j}(u)$. When

$p_{\gamma_j}(u) > 1$, the obervation with rank u is over-sampled, and when $p_{\gamma_j}(u) < 1$ the observation with rank u is under-sampled. The density of the distribution in group j then verifies:

$$f_j(y) = \frac{p_{\gamma_j}\left(\widetilde{F}(y)\right)\widetilde{f}(y)}{\int_{\underline{y}_j}^{\overline{y}_j} p_{\gamma_j}\left(\widetilde{F}(z)\right)\widetilde{f}(z)\,dz} \tag{6}$$

where $\widetilde{f}(y)$ is the baseline distribution and $\underline{y}_j$ is the minimum value and $\overline{y}_j$ the maximum in the group j. We have:

$$F_j(y) = \frac{\int_{\underline{y}_j}^{y} p_{\gamma_j}\left(\widetilde{F}(z)\right)\widetilde{f}(z)\,dz}{\int_{\underline{y}_j}^{\overline{y}_j} p_{\gamma_j}\left(\widetilde{F}(z)\right)\widetilde{f}(z)\,dz} = \frac{\int_{\underline{u}_j}^{\tilde{F}(y)} p_{\gamma_j}(v)\,dv}{\int_{\underline{u}_j}^{\overline{u}_j} p_{\gamma_j}(v)\,dv}$$

where $\underline{u}_j = \tilde{F}\left(\underline{y}_j\right)$ and $\overline{u}_j = \tilde{F}\left(\overline{y}_j\right)$. $\underline{u}_j$ (resp. $\overline{u}_j$) is the lowest (resp. highest) rank of the group j observations in the baseline distribution. We have $F_j(y) = r_{\gamma_j}\left(\tilde{F}(y)\right)$ where r_{γ_j} is the increasing function defined by:

$$r_{\gamma_j}(u) = \frac{P_{\gamma_j}(u)}{P_{\gamma_j}(\overline{u}_j)}, \text{ for } u \in \left[\underline{u}_j, \overline{u}_j\right] \tag{7}$$

where $P_{\gamma_j}(u) = \int_{\underline{u}_j}^{u} p_{\gamma_j}(v)\,dv$. Note that $z_{\beta_j}(y) = y$. We get:

$$r_{\gamma_j}^{-1}(v) = P_{\gamma_j}^{-1}\left(P_{\gamma_j}(\overline{u}_j)\,v\right), \text{ for } v \in [0,1] \tag{8}$$

We want to compose $r_{\gamma_1}^{-1}(v)$ with the function r_{γ_2} for admissible value of v. We have $r_{\gamma_1}^{-1}(v) \in [\underline{u}_1, \overline{u}_1]$ for $v \in [0,1]$. Hence, $r_{\gamma_0,2} = r_{\gamma_2} \circ r_{\gamma_1}^{-1}(v)$ is well defined only for $r_{\gamma_1}^{-1}(v) \in [\max(\underline{u}_1, \underline{u}_2), \min(\overline{u}_1, \overline{u}_2)]$, that is to say for $v \in \left[\max\left(0, r_{\gamma_1} \circ r_{\gamma_2}^{-1}(0)\right), \min\left(1, r_{\gamma_1} \circ r_{\gamma_2}^{-1}(1)\right)\right]$. We finally get:

$$\begin{aligned} r_{\gamma_0,2}(y) &= r_{\gamma_2}\left(P_{\gamma_1}^{-1}\left(P_{\gamma_1}(\overline{u}_1)\,y\right)\right) = \frac{P_{\gamma_2}\left(P_{\gamma_1}^{-1}\left(P_{\gamma_1}(\overline{u}_1)\,y\right)\right)}{P_{\gamma_2}(\overline{u}_2)}, \\ \text{for } y &\in \left[\max\left(0, r_{\gamma_1} \circ r_{\gamma_2}^{-1}(0)\right), \min\left(1, r_{\gamma_1} \circ r_{\gamma_2}^{-1}(1)\right)\right] \end{aligned} \tag{9}$$

Note that (6) can be rewritten as: $f_j(y) = r'_{\gamma_j}\left(\widetilde{F}(y)\right)\widetilde{f}(y)$. Hence r'_{γ_j} is the sampling function that selects observations into the group j depending on their rank in the baseline distribution. It is possible to show that: $f_2(y) = r'_{\gamma_0}(F_1(y))\,f_1(y)$ and $f_1(y) = r_{\gamma_0}^{-1\prime}(F_2(y))\,f_2(y)$.[2] This means that r'_{γ_0} (resp. $r_{\gamma_0}^{-1\prime}$) can be interpreted as the sampling function that selects obervations from group 1 (resp. 2) into group 2 (resp. 1) depending on their rank in group 1 (resp. 2).

[2] Indeed, combining (6) for groups 1 and 2, we get: $f_2(x) = \frac{P_{\gamma_1}(1)}{P_{\gamma_2}(1)}\frac{p_{\gamma_2}\left(\widetilde{F}(x)\right)}{p_{\gamma_1}\left(\widetilde{F}(x)\right)} f_1(x)$. We also have from (9): $r'_{\gamma}(F_1(x)) = \frac{P_{\gamma_1}(1)}{P_{\gamma_2}(1)}\frac{p_{\gamma_2}\left(P_{\gamma_1}^{-1}(P_{\gamma_1}(1)F_1(x))\right)}{p_{\gamma_1}\left(P_{\gamma_1}^{-1}(P_{\gamma_1}(1)F_1(x))\right)}$. Using (8), we get $F_1(x) = r_{\gamma_1}\left(\tilde{F}(x)\right) \iff \tilde{F}(x) = r_{\gamma_1}^{-1}(F_1(x)) = P_{\gamma_j}^{-1}\left(P_{\gamma_j}(1)\,F_1(x)\right)$ we get the result. The same line of arguments hold for group 2.

Consider the special case of *smooth random selection* in which $p_{\gamma_j}(u) \in [0,1]$ for all $u \in [0,1]$ with p_{γ_j} differentiable, and more specifically the fonction $p_{\gamma_j}(u) = (1-u)^{\gamma_j}$ for $\gamma_j > 0$ (selection is here decreasing in the rank in the distribution). We then get: $P_{\gamma_j}(u) = \frac{1-(1-u)^{\gamma_j+1}}{\gamma_j+1}$, thus $r_{\gamma_j}(u) = 1-(1-u)^{\gamma_j+1}$. Finally, $r_{\gamma_0,2}(y) = r_{\gamma_2} \circ r_{\gamma_1}^{-1}(y) = 1-(1-y)^{\frac{\gamma_2+1}{\gamma_1+1}}$ and only the combination of parameters $\frac{\gamma_2+1}{\gamma_1+1} = \gamma_0+1$ is identified. When $\gamma_0 < 0$ (resp. $\gamma_0 > 0$), observations in group 2 are more (resp. less) selected in the top of the distribution than observations in group 1.

Consider now the special case of *deterministic selection* where there are both a left and a right censorship. In that case, we have: $p_{\gamma_j}(u) = 1$ for $u \in \left[\underline{u}_j, \overline{u}_j\right] \subset [0,1]$ with $\gamma_j = \left(\underline{u}_j, \overline{u}_j\right)'$ and $p_{\gamma_j}(u) = 0$ for $u \notin \left[\underline{u}_j, \overline{u}_j\right]$. We get: $P_{\gamma_j}(u) = \min\left(u, \overline{u}_j\right) - \min\left(u, \underline{u}_j\right)$, and thus $r_{\gamma_j}(u) = \frac{\min(u,\overline{u}_j)-\min\left(u,\underline{u}_j\right)}{\overline{u}_j-\underline{u}_j}$. Hence, $r_{\gamma_j}^{-1}(v) = \left(\overline{u}_j - \underline{u}_j\right)v + \underline{u}_j$ for $v \in [0,1]$, and $r_\gamma(y) = \frac{\underline{u}_1-\underline{u}_2}{\overline{u}_2-\underline{u}_2} + \frac{\overline{u}_1-\underline{u}_1}{\overline{u}_2-\underline{u}_2}y$, for $y \in \left[\max\left(0, \frac{\underline{u}_2-\underline{u}_1}{\overline{u}_1-\underline{u}_1}\right), \min\left(1, \frac{\overline{u}_2-\underline{u}_1}{\overline{u}_1-\underline{u}_1}\right)\right]$. Hence, only two combinations of the parameters, say $\underline{v} = \frac{\underline{u}_1-\underline{u}_2}{\overline{u}_2-\underline{u}_2}$ and $\overline{v} = \frac{\overline{u}_1-\underline{u}_2}{\overline{u}_2-\underline{u}_2}$, are identified on the common supportdefined by $\left[\max\left(0, \frac{-\underline{v}}{\overline{v}-\underline{v}}\right), \min\left(1, \frac{1-\underline{v}}{\overline{v}-\underline{v}}\right)\right]$. Note that we have $r_\gamma(y) = \underline{v} + \left(\overline{v} - \underline{v}\right)y$: $\underline{v}$ (resp $\overline{v}$) is the relative left-censorship (resp. right-censorship) truncation parameter of the distribution 2 with respect to distribution 1. When $\underline{v} < 0$ (resp $\overline{v} > 1$), distribution 1 is more left-censored (resp. right-censored) than distribution 2.

4 Estimation

The parameters of the model are identified from equations (2) which are equalities between transformed quantiles of the two groups. However, in practice, quantiles are not observed and need to be estimated. For the estimated quantiles, equations (2) are not verified anymore because of the sampling error. In this section, we explain how to recover the parameters of the model from the minimization of a criterium which is specified as a function of estimated quantiles. We derive the asymptotic law followed by the estimated parameters and provide some asymptotic results on this criterium.

4.1 Minimization criterium

We choose the minimization criterium to be the quadratic distance between the transformed quantiles of groups 1 and 2 on their common support. Consider the distribution in group 1 truncated below by $\underline{u}_{\gamma_0,1}$ and above by $\overline{u}_{\gamma_0,1}$. Any rank t in this truncated distribution corresponds to a rank $u_{\gamma_0,1}(t) = \underline{u}_{\gamma_0,1} + \left(\overline{u}_{\gamma_0,1} - \underline{u}_{\gamma_0,1}\right)t$ in distribution 1 to which is associated the quantile $\underline{\lambda}_{\gamma_0,1}(t) = \lambda_1\left(u_{\gamma_0,1}(t)\right)$ and the transformed rank in distribution 2: $\underline{r}_{\gamma_0,2}(t) = r_{\gamma_0,2}\left(u_{\gamma_0,1}(t)\right)$. Equations (2) can be rewritten as:

$$z_{\beta_0,2}\left(\underline{\lambda}_{\gamma_0,1}(t)\right) = \lambda_2\left(\underline{r}_{\gamma_0,2}(t)\right) \text{ for } t \in [0,1] \tag{10}$$

A minimization criterium associated to (2) is then defined as:

$$\widehat{C}_1(\beta,\gamma) = \int_0^1 w_1(t)\left[z_{\beta,2}\left(\widehat{\underline{\lambda}}_{\gamma_0,1}(t)\right) - \widehat{\lambda}_2\left(\underline{r}_{\gamma_0,2}(t)\right)\right]^2 \mathrm{d}t \tag{11}$$

where $w_1(.)$ is any strictly positive weight function, $\widehat{\lambda}_j$ is an estimator of λ_j, and $\widehat{\underline{\lambda}}_{\gamma_0,1}(t) = \widehat{\lambda}_1\left(u_{\gamma_0,1}(t)\right)$. The literature proposes many estimators of λ_j which can be summarized in a nesting specification given by Cheng and Parzen (1997):

$$\widehat{\lambda}_j(u) = \int_0^1 \widetilde{\lambda}_j(t)\, d_t K_n(u,t) \tag{12}$$

where $\widetilde{\lambda}_j$ is the sample quantile function and for each u, $K_n(u,\cdot)$ is a cdf on $[0,1]$. When $K_n(u,\cdot)$ is a mass point 1 at u, (12) gives the sample quantile function. When $K_n(u,t) = \frac{1}{h_n}k\left(\frac{t-u}{h_n}\right)$ where k is a density and h_n a sequence of reals, (12) gives the kernel estimator. We make an additional assumption to ensure that the minimization criterium has a solution:

A4: Φ is compact.

In fact, we can also use the equations (3) to get a minimization criterium. Consider the distribution in group 2 truncated below by $\underline{u}_{\gamma_0,2}$ and above by $\overline{u}_{\gamma_0,2}$. Any rank t in this truncated distribution corresponds to a rank $u_{\gamma_0,2}(t) = \underline{u}_{\gamma_0,2} + \left(\overline{u}_{\gamma_0,2} - \underline{u}_{\gamma_0,2}\right)t$ in distribution 2 to which is associated the quantile $\underline{\lambda}_{\gamma_0,2}(t) = \lambda_2\left(u_{\gamma_0,2}(t)\right)$ and the transformed rank in distribution 1: $\underline{r}_{\gamma_0,1}(t) = r_{\gamma_0,1}\left(u_{\gamma_0,2}(t)\right)$. Equations (3) can be rewritten as:

$$\lambda_1\left(\underline{r}_{\gamma_0,1}(t)\right) = z_{\beta_0,1}\left(\underline{\lambda}_{\gamma_0,2}(t)\right), \text{ for } t \in [0,1]$$

A minimization criterium associated to (3) is:

$$\widehat{C}_2(\beta,\gamma) = \int_0^1 w_2(t)\left[\widehat{\lambda}_1\left(\underline{r}_{\gamma_0,1}(t)\right) - z_{\beta,1}\left(\widehat{\underline{\lambda}}_{\gamma_0,2}(t)\right)\right]^2 \mathrm{d}t \tag{13}$$

where $w_2(.)$ is any strictly positive weight function, and $\widehat{\underline{\lambda}}_{\gamma_0,2}(t) = \widehat{\lambda}_2\left(u_{\gamma_0,2}(t)\right)$. The expression of $\widehat{C}_2$ is the same as that of $\widehat{C}_1$, switching groups 1 and 2. It is then possible to combine the two criteria to obtain a criterium where groups 1 and 2 are treated symetrically:

$$\widehat{C}(\beta,\gamma) = \widehat{C}_1(\beta,\gamma) + \widehat{C}_2(\beta,\gamma) \tag{14}$$

Note that the criterium (14) embeds both the criteria (11) and (13) that can be recovered respectively fixing $w_1(t) = 0$ and $w_2(t) = 0$ for all $t \in [0,1]$.

4.2 Asymptotics

We now study the asymptotic properties of the estimators of the parameters and the criterium. We make an additional assumptions:

A5a: for any function g, we have: $\sup_{u\in[0,1]}\left|\int_0^1 g(t)\, d_t K_n(u,t) - g(u)\right| \xrightarrow{P} 0$.
A5b: we have: $\left|\frac{\partial K_n}{\partial t}(u,t)\right| \leqslant M_0(u,n)$ for almost all t, with $\sup_{u\in[0,1]} M_0(u,n) = o\left(n^{\theta_0}\right)$, $\theta_0 < 1/2$.

Applying A5a to the function $\widetilde{\lambda}_j$ and using A5b ensures that the quantile estimator converges uniformely to its true value (see appendix). Note that A5a and A5b hold trivially when $K_n(u,\cdot)$ is a mass point 1 at u. When K_n is a kernel, A5a usually holds when the bandwidth tends to zero as n tends to infinity, and A5b holds if the convergence speed of the bandwidth is not too large. We then have the following consistency theorem for the estimated parameters:

Theorem 1 *Under A1-A5b, we have:* $\left(\widehat{\beta},\widehat{\gamma}\right) \xrightarrow{P} (\beta_0,\gamma_0)$ *the true value of the parameters.*
Proof. *See appendix.* ■

We now establish the asymptotic distribution of the estimated parameters. We denote $\underline{f}_{\gamma,j}\left(\underline{\lambda}_{\gamma,j}(t)\right) = \frac{f_j\left(\underline{\lambda}_{\gamma,j}(t)\right)}{\overline{u}_{\gamma,j}-\underline{u}_{\gamma,j}} 1_{t\in[0,1]}$ the density of distribution j truncated below by $\underline{u}_{\gamma,j}$ and above by $\overline{u}_{\gamma,j}$, where f_j is the density in group j. We also introduce N_j the number of observations in group j and $\widehat{N}_{\gamma,j}$ the number of uncensored observations (which are the observations corresponding to ranks between $\underline{u}_{\gamma,j}$ and $\overline{u}_{\gamma,j}$). We also introduce the two functions $\mu_1(t) = z_{\beta,2} \circ \underline{\lambda}_{\gamma,1}(t)$ and $\mu_2(t) = z_{\beta,1} \circ \underline{\lambda}_{\gamma,2}(t)$ as well as the two following vectors that will be used to write our convergence theorem:

$$
\begin{aligned}
V_{(\beta,\gamma),1}(t) &= 2\begin{pmatrix} \left.\frac{\partial z_{\beta,2}}{\partial \beta}\right|_{\underline{\lambda}_{\gamma,1}(t)} \\ -\left.\frac{\partial r_{\gamma,2}}{\partial \gamma}\right|_{u_{\gamma,1}(t)} \lambda_2' \circ \underline{r}_{\gamma,2}(t) \end{pmatrix} \\
V_{(\beta,\gamma),2}(t) &= 2\begin{pmatrix} \left.\frac{\partial z_{\beta,1}}{\partial \beta}\right|_{\underline{\lambda}_{\gamma,2}(t)} \\ -\left.\frac{\partial r_{\gamma,1}}{\partial \gamma}\right|_{u_{\gamma,2}(t)} \lambda_1' \circ \underline{r}_{\gamma,1}(t) \end{pmatrix}
\end{aligned}
$$

In the next theorem, we show that the estimated parameters converge in distribution to a normal law at speed $\sqrt{\frac{1}{\widehat{N}_{\gamma,1}}+\frac{1}{\widehat{N}_{\gamma,2}}}$ under some additional assumptions. The line of the proof is quite similar to that used to establish the asymptotic law of the maximum likelihood estimator. We first want to apply the mean value theorem to the first-order derivative of the $\widehat{C}$. For that purpose, we need $\widehat{C}$ to be twice differentiable. We thus make the assumptions:

A5c: $\frac{\partial K_n}{\partial t}$ *is twice differentiable in* u *and* $\left|\frac{\partial^{k+1}}{\partial u^k \partial t} K_n(u,t)\right| \leqslant M_k(u,n)$ *for almost all* t *and* $k \in \{1,2\}$, *with* $\sup\limits_{u\in[0,1]} M_k(u,n) = o\left(n^{\theta_k}\right)$, $\theta_k < 1/2$.
A5d: for $k \in \{1,2\}$ *and any function* g *that is twice differentiable, we have:* $\sup\limits_{u\in[0,1]} \left|\int_0^1 g(t) \frac{\partial^{k+1}}{\partial u^k \partial t} K_n(u,t)\, dt - \frac{\partial^k g}{\partial u^k}(u)\right| \xrightarrow{P} 0$.
A6: there is a lower bound $a > 0$ *such that* $f(x) \geqslant a$ *nearly everywhere.*

Assumption A5c ensures that the estimated quantiles $\widehat{\lambda}_j$, $j \in \{1,2\}$ which intervene in $\widehat{C}$ are twice differentiable, and so is $\widehat{C}$. Note that this property is not verified when $\widehat{\lambda}_j$ is the sample quantile estimator $\widetilde{\lambda}_j$. According to the mean value theorem, we have:

$$\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix} = -\left(\left.\frac{\partial^2 \widehat{C}}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\overline{\beta},\overline{\gamma})} \right)^{-1} \left.\frac{\partial \widehat{C}}{\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)}$$

where $\overline{\beta}$ lies between β_0 and $\widehat{\beta}$, and $\overline{\gamma}$ lies between γ_0 and $\widehat{\gamma}$. Assumption A5d ensures that the derivatives of quantiles $\widehat{\lambda}_j$ converge to the derivatives of quantiles λ_j. When K_n is a kernel, the assumption is verified if the kernel is sufficiently smooth and the bandwidth tends to zero fast enough when n tends to infinity (see Härdle, 1990). Finally, assumption A6 ensures that λ_j and its first and second-order derivatives to be lipschitzienne. This is because the derivatives of the quantiles functions can be bounded up to the third-order as they write as ratios such that their denominator stays away from zero. It is then possibility to show that the first right-hand side term converges to a given matrix (see Lemma 5 in appendix). This matrix would correspond to the Fisher information matrix for the maximum likelihood estimator. We can then establish that $\left.\frac{\partial \widehat{C}}{\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)}$ follows a normal distribution. In fact, we have that following asymptotic equivalence (see appendix):

$$\begin{aligned} \left.\frac{\partial \widehat{C}}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)} \sim & \int_0^1 w_1(t)\, V_{(\beta_0,\gamma_0),1}(t) \left[z_{\beta_0,2} \circ \underline{\widehat{\lambda}}_{\gamma_0,1}(t) - \widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t) \right] \mathrm{d}t \\ & + \int_0^1 w_2(t)\, V_{(\beta_0,\gamma_0),2}(t) \left[z_{\beta_0,1} \circ \underline{\widehat{\lambda}}_{\gamma_0,2}(t) - \widehat{\lambda}_1 \circ \underline{r}_{\gamma_0,1}(t) \right] \mathrm{d}t \end{aligned}$$

Terms in the integrals on the right-hand side are random because we use the estimated quantiles instead of the true value of the quantiles. In fact, these integrals write as the sum of some estimated quantiles. These estimated quantiles are not independent. Indeed, for any given $j \in \{1,2\}$, the quantiles $\widehat{\lambda}_j(u_1)$ and $\widehat{\lambda}_j(u_2)$ are correlated for any given u_1,$u_2 \in [0,1]^2$. In our asymptotic derivations, we thus need the joint asymptotic law of the estimated quantiles. We restrict our attention to kernel estimators and suppose that:

A7: the functions $K_n(\cdot,\cdot)$ are some kernel functions such that $K_n(u,t) = \frac{1}{h_n} k\left(\frac{t-u}{h_n}\right)$ where $0 < h_n \to 0$, k has a bounded support, $\int k(x)\,dx = 1$, $\widetilde{\lambda}$, $r_{\gamma_j}^{-1}$, and z_{β_j} have bounded first-order derivatives.

Under assumptions A1-A7, using Theorem 1.3 p429 in Falk (1985), we get that $\sqrt{N_j} f_j(\lambda)(\widehat{\lambda}_j - \lambda_j)$ converges to a standard brownian bridge. This property is enough to establish the result.

We now give our theorem which establishes the asymptotic distribution of the estimated parameters:

Theorem 2 *Under A1-A7, we have:*

$$\left(\frac{1}{\widehat{N}_{\gamma,1}} + \frac{1}{\widehat{N}_{\gamma,2}} \right)^{-\frac{1}{2}} \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix} \overset{d}{\Longrightarrow} N\left(0, \Gamma^{-1}\Omega\Gamma^{-1}\right)$$

where:

$$\begin{aligned} \Gamma &= \frac{1}{2}\int_0^1 \left[w_1(t)\, V_{(\beta_0,\gamma_0),1}(t)\, V^T_{(\beta_0,\gamma_0),1}(t) + w_2(t)\, V_{(\beta_0,\gamma_0),2}(t)\, V^T_{(\beta_0,\gamma_0),2}(t) \right] \mathrm{d}t \\ \Omega &= \int_0^1 T\{g\}(u)\, T\{g\}(u)'\, \mathrm{d}u \end{aligned}$$

with:

$$T\{g\}(u) = \int_u^1 g(t)\,\mathrm{d}t - \int_0^1 tg(t)\,\mathrm{d}t$$

where: $g(t) = \left[w_1(t)\,V_{(\beta_0,\gamma_0),1}(t) + w_2(s(t))\,V_{(\beta_0,\gamma_0),2}(s(t))\,z_{\beta_0,1}[\mu_1(t)]\right]\mu_1'(t)$

and: $s(t) = \frac{\underline{r}_{\gamma_0,2}(t) - \underline{u}_{\gamma_0,2}}{\overline{u}_{\gamma_0,2} - \underline{u}_{\gamma_0,2}}$

Proof. *See appendix.* ■

As the asymptotic law is quite intricate, confidence intervals may be computed by bootstrap. We now establish the asymptotic distribution of the minimization criterium (14) at the optimum. This asymptotic distribution can then be used to test the model. We have the following theorem:

Theorem 3 *Under A1-A7, we have:*

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}} + \frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1} \widehat{C}(\beta_0,\gamma_0) \overset{d}{\Longrightarrow} \int_0^1 w_1(t)\,\mu_1'(t)^2\,B_1(t)^2\,\mathrm{d}t + \int_0^1 w_2(s)\,\mu_2'(s)^2\,B_2(s)^2\,\mathrm{d}s \tag{15}$$

where $\mu_1(t) = z_{\beta_0,2}\left(\underline{\lambda}_{\gamma_0,1}(t)\right)$*,* $\mu_2(s) = \lambda_1\left(\underline{r}_{\gamma_0,1}(s)\right)$*,* $\widehat{N}_{\gamma_0,j}$ *is the number of uncensored observations in group* j*,* B_1 *and* B_2 *are some brownian bridges such that for all* $s,t \in [0,1]$*,* $B_1(t) = B_2(s)$ *with* s *and* t *such that:* $u_{\gamma_0,2}(s) = \underline{r}_{\gamma_0,2}(t)$*. We also have:*

$$\left(\frac{1}{\widehat{N}_{\widehat{\gamma},1}} + \frac{1}{\widehat{N}_{\widehat{\gamma},2}}\right)^{-1} \widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right) - \left(\frac{1}{\widehat{N}_{\gamma_0,1}} + \frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1} \widehat{C}(\beta_0,\gamma_0) \overset{d}{\sim} -\frac{1}{2}\sum_k D_{(\beta_0,\gamma_0),k} X_k^2 \tag{16}$$

where X_k*,* $k = 1,...,K$*, are some independent normal laws and* $D_{(\beta_0,\gamma_0),k}$ *is the* k^{th} *(positive) eigenvalue of* $\Lambda'\Gamma\Lambda$ *where* $\Gamma^{-1}\Omega\Gamma^{-1} = \Lambda\Lambda'$ *(which is a Choleski decomposition).*

Proof. See Appendix. ■

Theorem (3) establishes the asymptotic law of the criterium and thus provides a way to test of a the model. Indeed, the criterium should not be different from zero if the model is true and its value should thus be in the 95% confidence interval of the asymtpotic law.

The asymptotic law of the criterium can be decomposed in two terms. The first one (15) is the integral of a single weighted brownian bridge as it rewrites:

$$\begin{aligned} & \int_0^1 w_1(t)\,\mu_1'(t)^2\,B_1(t)^2\,\mathrm{d}t + \int_0^1 w_2(s)\,\mu_2'(s)^2\,B_2(s)^2\,\mathrm{d}s \\ = & \int_0^1 \left[w_1(t) + w_2(s(t))\,z_{\beta_0,1}'[\mu_1(t)]^2/s'(t)\right]\mu_1'(t)^2\,B_1^2(t)\,\mathrm{d}t \end{aligned} \tag{17}$$

This is obtained with a change in variable (see appendix). It comes from the fact that B_1 and B_2 are linked and correspond to the same brownian bridge. The second term of the asymptotic law (16) is a sum of weighted chi-squares (X_k). Note that the two terms (15) and (16) are not independent since X_k, $k = 1,...,K$ can be expressed as functions of B_1 and B_2.

Theorem (3) also shows that the asymptotic law of the criterium $\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)$ depends on the true value of the parameters (β_0,γ_0). This makes impossible to test that $\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)=0$. However, for some specific weights $w_1(t)=w_2(s(t))=\frac{1}{\left[1+z'_{\beta_0,1}[\mu_1(t)]^2/s'(t)\right]\mu'_1(t)^2}$, the integral (17) then writes as a pivotal statistic and follows the Van-Mises law $\int_0^1 B_1^2(t)\,\mathrm{d}t$ which has been tabulated in the literature (see Knott, 1974). By definition, we have: $\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)<\widehat{C}(\beta_0,\gamma_0)$. Denote C_a the asymptotic law of $\left(\frac{1}{\widehat{N}_{\widehat{\gamma},1}}+\frac{1}{\widehat{N}_{\widehat{\gamma},2}}\right)^{-1}\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)$. If the model is true, we should have:

$$H_0\colon\ C_a<\int_0^1 B_1^2(t)\,\mathrm{d}t$$

Testing that the model is true can be done by testing H_0. As H_0 only corresponds to an inequality between two distributions (and not an equality), the test might be not very powerful.

Note that the weights leading to the Cramer-Von Mises statistic depend on the true value of the parameters. A way to construct the CVM statistic is in two stages (as in GMM). A first set of parameters is estimated without weights. Then weights are computed using these estimators. A new criterium is then constructed from these weights and minimized. The value of the criterium at its optimum can be used to test H_0.

5 Implementation

5.1 Estimation of the parameters

We now explain how the parameters of the model can be estimated in practice. The theoretical model gives the specification of $z_{\beta,j}$ and $r_{\gamma,j}$ which forms are known. From these functions, it is possible to calculate the functions $\frac{\partial r_{\gamma,j}}{\partial\gamma}$ and $\frac{\partial z_{\beta,j}}{\partial\beta}$. The data consist in a set of values x_k^j for each group $j\in\{1,2\}$ and for $k\in\{1,...,N_j\}$. We order them in ascending order such that $y_1^j<...<y_{N_j}^j$. We then have some estimators of the quantiles corresponding to the ranks in the two groups: $\widehat{\lambda}_j\left(\frac{k}{N_j}\right)=y_k^j$.

For our minimization program, we need to compute the criterium for any admissible parameters $(\beta,\gamma)\in\Phi$. In particular, a couple is not admissible if $\underline{u}_{\gamma,1}>\overline{u}_{\gamma,1}$ or $\underline{u}_{\gamma,2}>\overline{u}_{\gamma,2}$. It is easy to compute $\underline{u}_{\gamma,1},\overline{u}_{\gamma,1},\underline{u}_{\gamma,2}$ and $\overline{u}_{\gamma,2}$ for any given values of the parameters, as the functions $r_{\gamma,j}$ are known. This allows us to discard parameters which are not admissible.

For admissible parameters, we now explain how to compute the criterium. We first rewrite its components as:

$$\begin{aligned}\widehat{C}_1(\beta,\gamma)&=\frac{1}{\overline{u}_{\gamma,1}-\underline{u}_{\gamma,1}}\int_{\underline{u}_{\gamma,1}}^{\overline{u}_{\gamma,1}}w_1\left(\frac{u-\underline{u}_{\gamma,1}}{\overline{u}_{\gamma,1}-\underline{u}_{\gamma,1}}\right)\left[z_{\beta,2}\left(\widehat{\lambda}_1(u)\right)-\widehat{\lambda}_2(r_{\gamma,2}(u))\right]^2\mathrm{d}u\\ \widehat{C}_2(\beta,\gamma)&=\frac{1}{\overline{u}_{\gamma,2}-\underline{u}_{\gamma,2}}\int_{\underline{u}_{\gamma,2}}^{\overline{u}_{\gamma,2}}w_2\left(\frac{u-\underline{u}_{\gamma,2}}{\overline{u}_{\gamma,2}-\underline{u}_{\gamma,2}}\right)\left[\widehat{\lambda}_1(r_{\gamma,1}(u))-z_{\beta,1}\left(\widehat{\lambda}_2(u)\right)\right]^2\mathrm{d}u\end{aligned}$$

To approximate the two integrals, we need to compute the differences between transformed quantiles $z_{\beta,2}\left(\widehat{\lambda}_1(u)\right)-\widehat{\lambda}_2(r_{\gamma,2}(u))$ for some $u\in\left[\underline{u}_{\gamma,1},\overline{u}_{\gamma,1}\right]$, or $\widehat{\lambda}_1(r_{\gamma,1}(v))-z_{\beta,1}\left(\widehat{\lambda}_2(v)\right)$ for some $v\in\left[\underline{u}_{\gamma,2},\overline{u}_{\gamma,2}\right]$. We can evaluate

$\widehat{\lambda}_1(u)$ for the group-1 observed ranks in $\left[\underline{u}_{\gamma,1},\overline{u}_{\gamma,1}\right]$ but we do not have any empirical counterpart for $\widehat{\lambda}_2(r_{\gamma,2}(u))$, since these ranks once transformed by $r_{\gamma,2}$ have no reason to match group-2 ranks. Hence, we need to estimate $\widehat{\lambda}_2(r_{\gamma,2}(u))$. Conversely, we can evaluate $\widehat{\lambda}_2(v)$ for the group-2 observed ranks in $\left[\underline{u}_{\gamma,2},\overline{u}_{\gamma,2}\right]$, but we do not have any empirical counterpart for $\widehat{\lambda}_1(r_{\gamma,1}(v))$ as these ranks once transformed by $r_{\gamma,1}$ have no reason to match ranks in group 1. Hence, we also need to estimate $\widehat{\lambda}_1(r_{\gamma,1}(v))$.

In practice, we can compute $\widehat{\lambda}_1\left(\frac{k}{N_1}\right)$ for $\frac{k}{N_1}\in\left[\underline{u}_{\gamma,1},\overline{u}_{\gamma,1}\right]$, and $\widehat{\lambda}_2\left(\frac{k}{N_2}\right)$ for all $\frac{k}{N_2}\in\left[\underline{u}_{\gamma,2},\overline{u}_{\gamma,2}\right]$. Hence, we construct estimators of the missing ranks in each group the following way:

- For $\frac{k}{N_1}\in\left[\underline{u}_{\gamma,1},\overline{u}_{\gamma,1}\right]$, denote $k^*\in\{1,...,N_2\}$ the integer such that: $\frac{k^*}{N_2}<r_{\gamma,2}\left(\frac{k}{N_1}\right)<\frac{k^*+1}{N_2}$. The estimator of $\lambda_2\left(r_{\gamma,2}\left(\frac{k}{N_1}\right)\right)$ is defined as:

$$\widehat{\lambda}_2\left(r_{\gamma,2}\left(\frac{k}{N_1}\right)\right)=\frac{\left(\frac{k^*+1}{N_2}-r_{\gamma,2}\left(\frac{k}{N_1}\right)\right)\widehat{\lambda}_2\left(\frac{k^*+1}{N_2}\right)+\left(r_{\gamma,2}\left(\frac{k}{N_1}\right)-\frac{k^*}{N_2}\right)\widehat{\lambda}_2\left(\frac{k^*}{N_2}\right)}{\frac{1}{N_2}}$$

- For $\frac{k}{N_2}\in\left[\underline{u}_{\gamma,2},\overline{u}_{\gamma,2}\right]$, denote $k^*\in\{1,...,N_1\}$ the integer such that: $\frac{k^*}{N_1}<r_{\gamma,1}\left(\frac{k}{N_2}\right)<\frac{k^*+1}{N_1}$. The estimator of $\lambda_1\left(r_{\gamma,1}\left(\frac{k}{N_2}\right)\right)$ is defined as:

$$\widehat{\lambda}_1\left(r_{\gamma,1}\left(\frac{k}{N_2}\right)\right)=\frac{\left(\frac{k^*+1}{N_1}-r_{\gamma,1}\left(\frac{k}{N_2}\right)\right)\widehat{\lambda}_1\left(\frac{k^*+1}{N_1}\right)+\left(r_{\gamma,1}\left(\frac{k}{N_2}\right)-\frac{k^*}{N_1}\right)\widehat{\lambda}_2\left(\frac{k^*}{N_1}\right)}{\frac{1}{N_1}}$$

We define the vector V_j of the ranks in group j, observed or constructed, sorted in ascending sequence. The total number of possible ranks in V_1 and V_2 is the same and equals the total number of observations in the two groups $\widehat{N}_\gamma=\widehat{N}_{\gamma,1}+\widehat{N}_{\gamma,2}$. Note that for each $i\in\left\{1,...,\widehat{N}_\gamma\right\}$, we have $V_2[i]=r_{\gamma,2}(V_1[i])$. The two components of the criterium are estimated by:

$$\widehat{C}_1(\beta,\gamma)=\frac{1}{2}\sum_{i=2}^{\widehat{N}}\left[\begin{array}{c} w_1\left(\frac{V_1[i]-\underline{u}_1}{\overline{u}_1-\underline{u}_1}\right)\left[z_{\beta,2}\left(\widehat{\lambda}_1(V_1[i])\right)-\widehat{\lambda}_2(V_2[i])\right]^2 \\ +w_1\left(\frac{V_1[i-1]-\underline{u}_1}{\overline{u}_1-\underline{u}_1}\right)\left[z_{\beta,2}\left(\widehat{\lambda}_1(V_1[i-1])\right)-\widehat{\lambda}_2(V_2[i-1])\right]^2\end{array}\right]\frac{(V_1[i]-V_1[i-1])}{\overline{u}_1-\underline{u}_1} \quad (18)$$

$$\widehat{C}_2(\beta,\gamma)=\frac{1}{2}\sum_{i=2}^{\widehat{N}}\left[\begin{array}{c} w_2\left(\frac{V_2[i]-\underline{u}_2}{\overline{u}_2-\underline{u}_2}\right)\left[\widehat{\lambda}_1(V_1[i])-z_{\beta,1}\left(\widehat{\lambda}_2(V_2[i])\right)\right]^2 \\ +w_2\left(\frac{V_2[i-1]-\underline{u}_2}{\overline{u}_2-\underline{u}_2}\right)\left[\widehat{\lambda}_1(V_1[i-1])-z_{\beta,1}\left(\widehat{\lambda}_2(V_2[i-1])\right)\right]^2\end{array}\right]\frac{(V_2[i]-V_2[i-1])}{\overline{u}_2-\underline{u}_2} \quad (19)$$

The criterium cannot be minimized directly with usual optimization tools as the criterium is not continuous in γ (there are some jumps depending on the number of observations which are censored). Hence, the minimization procedure is decomposed into two stages. For a given admissible γ, it is possible to minimize the criterium with respect to β using some standard optimization tools as the criterium is usually continuous and derivable wih respect to β. Then a grid search on γ can be performed.

5.2 Computation of the test statistic

We now explain how to construct the test statistic $\left(\frac{1}{\widehat{N}_{\widehat{\gamma},1}}+\frac{1}{\widehat{N}_{\widehat{\gamma},2}}\right)^{-1}\widehat{C}_{\widehat{w}}\left(\widehat{\beta},\widehat{\gamma}\right)$, where C_w is the criterium corresponding to weights given by: $w_1(t)=w_2(s(t))=\frac{1}{\left[1+z'_{\beta_0,1}[\mu_1(t)]^2/s'(t)\right]\mu'_1(t)^2}$, $\widehat{\beta}$ and $\widehat{\gamma}$ are the consistent estimators of the parameters obtained when weights equal one (see the previous subsection for their estimation). The number of uncensored observations $\widehat{N}_{\widehat{\gamma},j}$ is the number of observations such that their rank is between $\overline{u}_{\widehat{\gamma},j}$ and $\underline{u}_{\widehat{\gamma},j}$. We need to compute some estimators of the weights. For than purpose, we need to get some estimators of $\mu_1(t)$, $\mu'_1(t)$ and $s'(t)$. In fact, we have:

$$\mu_1(t)=z_{\beta_0,2}\circ\underline{\lambda}_{\gamma_0,1}(t)=\lambda_2\left(\underline{r}_{\gamma_0,2}(t)\right)$$

Hence, an estimator of $\mu_1(t)$ is:

$$\widehat{\mu_1}(t)=\frac{1}{2}\left[z_{\widehat{\beta},2}\circ\widehat{\underline{\lambda}}_{\widehat{\gamma},1}(t)+\widehat{\lambda}_2\left(\underline{r}_{\widehat{\gamma},2}(t)\right)\right]$$

where the quantile function $\widehat{\lambda}_j$ are replaced by the estimators given in the previous subsection. We also have:

$$\begin{aligned}\mu'_1(t) &= \left(\overline{u}_{\gamma_0,1}-\underline{u}_{\gamma_0,1}\right)\lambda'_1\left(u_{\gamma_0,1}(t)\right)z'_{\beta_0,2}\circ\underline{\lambda}_{\gamma_0,1}(t)\\ &= \left(\overline{u}_{\gamma_0,2}-\underline{u}_{\gamma_0,2}\right)r'_{\gamma_0,2}\left(u_{\gamma_0,2}(t)\right)\lambda'_2\left(\underline{r}_{\gamma_0,2}(t)\right)\end{aligned}$$

Hence, an estimator of $\mu'_1(t)$ is:

$$\widehat{\mu}'_1(t)=\frac{1}{2}\left[\begin{array}{c}\left(\overline{u}_{\widehat{\gamma},1}-\underline{u}_{\widehat{\gamma},1}\right)\widehat{\lambda}'_1\left(u_{\widehat{\gamma},1}(t)\right)z'_{\widehat{\gamma},2}\circ\widehat{\underline{\lambda}}_{\widehat{\gamma},1}(t)\\ +\left(\overline{u}_{\widehat{\gamma},2}-\underline{u}_{\widehat{\gamma},2}\right)r'_{\widehat{\gamma},2}\left(u_{\widehat{\gamma},2}(t)\right)\widehat{\lambda}'_2\left(\underline{r}_{\widehat{\gamma},2}(t)\right)\end{array}\right]$$

where $\widehat{\lambda}'_j$ are some estimators of the quantile function derivatives constructed below. Finally, using the expression of $s(t)$, we can construct an estimator of its derivative:

$$\widehat{s}'(t)=\frac{\overline{u}_{\widehat{\gamma},1}-\underline{u}_{\widehat{\gamma},1}}{\overline{u}_{\widehat{\gamma},2}-\underline{u}_{\widehat{\gamma},2}}r'_{\widehat{\gamma},2}\left(u_{\widehat{\gamma},1}(t)\right)$$

The estimated weights are then:

$$\widehat{w}_1(t)=\widehat{w}_2(s(t))=\frac{1}{\left[1+z'_{\widehat{\beta},1}\left[\widehat{\mu}_1(t)\right]^2/\widehat{s}'(t)\right]\widehat{\mu}'_1(t)^2}$$

For any given $j\in\{1,2\}$, some estimators of the quantile function derivatives $\widehat{\lambda}'_j$ can be recovered from Siddiqui (1960), and Bloch and Gartswirth (1968). These estimators can be easily computed in the following way. Consider the two sets of ordered data: $y^j_1<...<y^j_{N_j}$ for $j\in\{1,2\}$. For a given u, the estimator writes:

$$\widehat{\lambda}'_j(u)=N_j\frac{y_{[k_1]}-y_{[k_0]}}{k_1-k_0}$$

where $k_1=[nu+m]$ (with $[\bullet]$ the integer function) if $nu+m<n$ and $k_1=n$ otherwise, $k_0=[nu-m]$ if $nu-m\geqslant 1$ and $k_0=1$ otherwise. Weights should be estimated for all the ranks used in the computation of the test statistic with formulas (18) and (19). This means that w_1 should be compute for the ranks $\frac{k}{N_1}$ for $k=1,...,N_1$ and $r_{\widehat{\gamma},1}\left(\frac{k}{N_2}\right)$ for $k=1,...,N_2$. Similarly, w_2 should be computed for the ranks $\frac{k}{N_2}$ for $k=1,...,N_2$ and $r_{\widehat{\gamma},2}\left(\frac{k}{N_1}\right)$ for $k=1,...,N_1$. We finally get the test statistic using (18) and (19).

6 Monte-Carlo simulations

We now present some Monte-Carlo simulations to assess the performances of our estimators. We consider two baseline distributions $\tilde{F}$ with different shapes: a normal and a pareto. The baseline distribution is restricted to a compact support to verify the assumptions of the model. We thus truncate a given percentage of the distribution on the left and on the right. In the baseline simulation, the truncation is set to be $\tau = 2.5\%$ on each side. The distribution in the first group is supposed to be equal to the baseline distribution: $\tilde{F} = F_1$. The distribution in the second group is derived from the baseline distribution with two transformations. The values are transformed with an homothecy: $z_\beta(x) = ax + b$ where $a = 1.2$ and $b = 1$ in the baseline simulation. The ranks are transformed with a selection process that can be of three different kind (see subection 3.2 for details):

1. Random selection: $r_\gamma(u) = 1 - (1-u)^{\gamma+1}$, $u \in [0,1]$, with $\gamma = 0.5$.
2. Left truncation: $r_\gamma(u) = \gamma + (1-\gamma)u$ for $u \in \left[\max\left(0, \frac{-\gamma}{1-\gamma}\right), 1\right]$. We want 10% of the distribution to be left-truncated. This corresponds to $\gamma = -0.11$.
3. Linear selection: $r_\gamma(u) = 2\gamma u - \gamma^2 u^2$ with $\gamma = 2$.

Data are generated in the following way. We consider that there are N_j observations in group j before any transformation takes place. In the baseline case, we fix $N_1 = N_2 = 10,000$.
For group 1, we draw some rank values u_{1i}, $i \in \{1,..,N_1\}$ in a uniform law $[\tau, 1-\tau]$ and take the inverse of the baseline cumulative to generate the values of our variable of interest: $y_{1i} = \widetilde{F}^{-1}(u_{1i})$.
For group 2, we draw some rank values u_{2i}, $i \in \{1,..,N_2\}$ in a uniform law $[\tau, 1-\tau]$ and assess whether the corresponding observations should be selected depending on the selection process:

1. Random selection: an observation i is selected with probability $(1-u_i)^\gamma$. Hence, we also draw a series of values v_{2i}, $i \in \{1,..,N_2\}$ in a uniform law $[0,1]$. Observation i is selected if $v_{2i} \leqslant (1-u_i)^\gamma$.
2. Left truncation: observation i is selected if $u_{2i} \geqslant \gamma$.
3. Linear selection: an observation i with a rank $u_{2i} \in \left[0, \frac{1}{\gamma}\right]$ is selected with probability $2\gamma(1-\gamma u_{2i})$. For this observation, we grow a value v_{2i} in a uniform law $[0,1]$ and keep the observation if $v_{2i} \leqslant 2\gamma(1-\gamma u_{2i})$. An observation i with rank $u_{2i} > \frac{1}{\gamma}$ is not selected and is thus discarded.

For observations still in sample 2 after selection, we generate the values of our variable of interest as $y_{2i} = a\widetilde{F}^{-1}(u_{2i}) + b$.
We then compute the criterium given by (14). In the baseline simulation, we consider that all weights are equal to one: $w_1(t) = 1$ and $w_2(t) = 1$ for all $t \in [0,1]$. All computation details are given in the previous section on the implementation.

We now present the results obtained with the Monte-Carlo simulations for 100 replications. For each coefficient a, b and γ, we report the estimated coefficient and the root mean-square error (RMSE). We first examine the case where the baseline distribution is normal. Table 1 shows that for random selection, parameters are not biased (column 1) and are very accurately estimated (col. 2). Indeed, the mean value of the translation parameter b (.991) is close to 1 and its RMSE is very small at .033. Similarly, the mean value of the dilatation parameter a (1.196) stands very close to 1.2 and its RMSE is very small at .014. Finally, the selection parameter γ is estimated still reasonably well with a mean of close to .5 (value .487) and a RMSE at .058. The test statistic computed with the true value of the parameters has a median of .103 and a RMSE of .078. As the 5% threshold is 0.461 (see Knott, 1974), the model is accepted for nearly all the iterations. When using the estimated parameters to compute the test statistic, the median drops to .037 and the $RMSE$ to .020. This suggests a discrepancy in the power of the test when using the estimated parameters instead of the true (unobserved) ones (see below for more on that).

We assess how the performances of the method vary with the number of observations. Of course, when this number gets smaller, estimators behave less well. For $N_1 = N_2 = 1000$, the estimated parameters are biased (col. 3). The bias is quite small for the translation and dilatation parameters (8% and 2% of the true value, respectively). It is larger and significant for the selection parameter (20% of the true value). Also, parameters are less precisely estimated since the RMSE are more than three times larger than in the baseline case for all the estimated parameters (col. 4). Conversely, the accuracy increases when the number of observations gets larger. When $N_1 = N_2 = 100,000$, the means of the estimated parameters are nearly equal to their true value (col. 5). The RMSE get very small (col. 6), that is 2.5 times less than in the baseline case.

We then assess how the estimators behave when we change the truncation parameter τ of the baseline distribution. When τ gets smaller, estimators are less precisely estimated as expected. This is because quantiles at extremes which enter the minimization criterium are estimated with less accuracy. For $\tau = 1\%$, the means of estimated parameters get slightly away from their true value (col. 7). The discrepancy is the largest for the selection parameters with a bias that reaches 8% of the true value. The RMSE of all the parameters are also more than 50% larger (col. 8). Biases get very large when $\tau = 0\%$. It even reaches 46% of the true value for the selection parameters (col. 9). In the same way, the RMSE get large (col. 10). Conversely, we also experimented what happens when the truncation parameter τ gets larger. The performances improve a bit compared to the baseline case: the means of estimated parameters are a bit closer from the true value (col. 11) and RMSE are smaller (col. 12).

Finally, as we are mostly interested in the selection process in this paper, we experimented what happens when the selection parameter varies. Results are still very good for smaller γ taking the values .2 and .05 (col. 13-16). Note that in all our alternative specifications (col. 3-16), the test statistic hopefully always accepts the model at a 5% level whether it is computed with the true or the estimated value of the parameters.

We then conduct the same kind of analysis for left truncation (Table 2). The method performs very well and give very accurate results for all alternative specifications. Interestingly, the bias when $N_1 = N_2 = 1,000$ is far smaller than in the random selection case. Moreover, when there is no truncation of the baseline distribution ($\tau = 0\%$), the difference between the means of estimated parameters and their true value is negligible. These results suggest that the method should be very robust for left truncation in the empirical applications (provided that the distributions in the two groups are close to normal).

We also apply the method in the linear selection case (Table 3). Results are not as good as in the previous cases. In the baseline specification, there is a significant bias on the estimated parameters. This bias goes up to 20% for the translation parameter. Also RMSE are very large. The method performs very poorly for $N_1 = N_2 = 1,000$ but give good results for $N_1 = N_2 = 100,000$. Interestingly, there does not seem to be a bias on the estimated parameters when $\tau = 0\%$ even if the RMSE are quite large. This suggests that the truncation of the baseline distribution to get a bounded support may distort too much the shape of the distribution to get reliable results. When using a linear selection in practice, it would thus be better not to truncate the baseline distribution (or only very mildly to avoid extreme values). Finally, the performances of the method are similar when changing the selection Parameter.

We then repeat the same simulation exercise for a Pareto baseline distribution. Results are quite similar to those obtained with a normal baseline in the three selection specifications, which suggests that the method is quite robust to the shape of the baseline distribution. Note however that for random selection (Table 4), the biases are larger in the Pareto case when not truncating (much) the baseline distribution ($\tau = 1\%$ or $\tau = 0\%$). This is not surprising as the Pareto distribution has some heavy tails and quantiles are more spread. Biases also appear for the Pareto baseline when the selection is a left truncation and $\tau = 0\%$ (Table 5), whereas it is not the case for the normal baseline. Finally, the linear selection performs quite poorly with the Pareto baseline (Table 6).

We finally tried to asss the power of our specification test. We generated the data with a Pareto (resp. normal) distribution with variance one, but supposed that the baseline was normal (resp. Pareto). As the normal and Pareto distributions are very different, we expected the specifications to be heavily rejected. In that this is the case for all specifications (Table 7 and 8).

7 Application

We now apply our method to study the wage difference between males and females. Indeed, there is a large literature on discrimination against females (see *XXX*). In particular, it is said that females have a very limited access to the best high-skilled jobs and their wage usually stays below a glass ceiling. We want to assess the relative importance of a uniform discrimination against all females and the glass ceiling effect. Most of the recent literature studying the glass ceiling uses wage quantile regressions (see Albrecht, Björklund and Vroman, 2003). It is said to be a glass ceiling if, after controlling for individual characteristics such as the diploma, the wage difference between males and females is larger, the higher the quantile. The job occupation is usually not included in the controls as the authors want to capture the barriers to the best high-skilled jobs in the wage difference. Our method departs from this approach as it does not assess the existence of a glass ceiling from the comparison between separate quantile regressions, but rather infer it from the whole shapes of the wage distributions of males and females. Indeed, it models the glass ceiling (and the uniform distribution) as some transformations between the two wage distributions and estimates the underlying parameters. The two transformations can be described independently in the following way:

- The glass ceiling makes the jobs at the top of the wage distribution less accessible for females than for males. It can be modelled stating that for a given wage ω, the relative chances for females of getting a wage higher than ω, $\frac{P(\omega_f>\omega)}{P(\omega_m>\omega)}$, decreases with ω. In fact, we model the relative chances such that:

$$\frac{P(\omega_f > \omega)}{P(\omega_m > \omega)} = [P(\omega_m > \omega)]^{\gamma} \tag{20}$$

  with $\gamma > 0$. This specification corresponds to the first example proposed for r_γ in section 3. Indeed, denoting $G_f(\omega) = G(\omega_f \leqslant \omega)$ and $G_m(\omega) = P(\omega_m \leqslant \omega)$, we have: $G_f(\omega) = 1-[1-G_m(\omega)]^{\gamma+1}$. Rewriting $G_m(\omega) = u$ and $r_\gamma(u) = G_f\left[G_m^{-1}(u)\right]$, we get: $r_\gamma(u) = 1-(1-u)^{\gamma+1}$ which is similar to the smooth selection case.

- For a job yielding a given observed wage ω_m for males, the uniform discrimination makes females earn only $\omega_f = \omega_m + \alpha$ with $\alpha < 0$.

Whereas the glass ceiling squeezes the wages obtained by females to the left compared to males, the uniform discrimination yields a uniform wage loss for all females. We want to estimate the squeezing and translation parameters, and test the fit of the model.

In our application, we use some administrative data on hourly wages in 2003 from the *Déclarations Annuelles des Salaires (DADS)*. These data are exhaustive for the private sector. We need to construct a sample of workers whose characteristics are homogenous, and who are old enough for gender differences in careers to have become significant. We choose to focus on full-time workers aged between 35 and 40. As we do not have any information on the diploma, we restrict our attention to white collars (including executives, engineers, managers and marketing staff). Hence, we will examine the glass ceiling only among skilled workers. Beside, our sample restriction allows us to avoid the issue of the minimum wage. The wage distributions of males and females are both trimmed by 1% at each tail to avoid reporting errors.

We select four subsamples corresponding to four differentiated two-digit sectors with enough observations to apply our method: equipment goods, bank and insurance, computers, and social sector workers. Table 9 reports some descriptive statistics for these subsamples. The proportion of females varies a lot across sectors. Whereas it is low for equipment goods (14%) and computers (19%), it goes far higher for bank and insurance (33%), and social sector workers (44%). Not surprisingly, the average hourly wage is the highest in the bank-and-insurance sector, followed by computers, equipment goods, and social sector (education and health). In all sectors, the average hourly wage is lower for females than for males. The mean wage difference is the highest in the bank-and-insurance sector (26% which correspond to 8.7 euros). In other sectors, the mean wage difference is much lower going down to 8.7% (1.7 euros) in the social sector, 5.7% (1.5 euros) for computers and finally 4.5% (1.0 euros) for equipment goods. In fact, wage differences between men and women occur for the whole shape of the wage distributions. Figures 1-4 show that for all sectors, the wage distributions of males and females are unimodal. However, for bank and insurance, females have a far higher peak than males and they are concentrated at the bottom of the overall wage distribution. Their distribution looks squeezed to the left and/or left-translated. The distributions in the social sector and for computers share the same pattern except that the peak for females is closer to the peak for males. Hence, curves for males and females look more similar. Finally, for equipment goods, the peak is slightly higher for males than for females. The wage distribution for females however still looks translated to the left compared to that of males. We then estimate the parameters of the model and bootstrap the standard errors and confidence intervals with replacement for $1,000$ replications. Results are reported in Table 10. They show that the translation parameter is negative and significant for all the four sectors. This suggests some uniform discrimination against females. The effect is the highest for computers (1.1 euros) and the lowest for equipment goods (0.5 euros). In fact, the union agreements applied in the sector of equipment goods grant workers within a socio-professional category some wages which are nearly equivalent. This may explain the lower discrimination against females. Such union agreements are also applied in the computer sector but jobs are heterogenous. Some give access to bonuses whereas others do not. There may be a selection process such that wages with bonuses are mostly attributed to males. This would give rise to discrimination.

Interestingly, the squeezing parameter is positive and significant for all the four sectors. It suggests a glass ceiling effect. The effect is the highest for bank and insurance, and the lowest for equipment goods. It is possible to construct some meaningful quantities from the squeezing parameter. Indeed, consider the wage ω_0 at the last decile of the distribution of males, that is the wage such that: $P\left(\omega_m > \omega_0\right) = 0.1$. Put differently, 10% of males get a wage higher than ω_0. The corresponding proportion for females is $P\left(\omega_f > \omega_0\right) = 0.1^{1+\gamma}$. This proportion ranges from 3.0% in the sector of bank and insurance (which is very small) to 7.6% in the sector of equipment goods (which is significantly below ten percent). Another way to quantify the squeezing is to estimate the glass ceiling effect on the wage of a female at some given quantiles of the wage distribution of males. In fact, rewriting (20), it is possible to show that, because of the glass ceiling, a female who would earn x was she a male is going to get $y = \lambda_m \left[1 - \left[1 - G_m\left(x\right)\right]^{\frac{1}{\gamma+1}}\right]$ where λ_m is the quantile function for males. We compute the wage loss $y - x$ at the first decile, the median and the last decile of the male wage distribution. This loss writes: $\lambda_m \left[1 - \left[1 - q\right]^{\frac{1}{\gamma+1}}\right] - \lambda_m\left(q\right)$ respectively for $q = 0.1$, 0.5 and 0.9. Results reported in Table 11 show that not surprisingly the wage loss at the first decile is smaller than the loss due to the uniform discrimination. This is because there is nearly no glass ceiling for lower wages. At the median, the glass ceiling effect is larger than the uniform discrimination effect only in two sectors (the social, and bank and insurance sectors). Finally, the effect of the glass ceiling is very important at the last decile in all sectors, except maybe equipment goods. In all sectors, it is much larger than the effect of the uniform discrimination. Whereas, females at the last decile incur a loss of 1.7 euros in the sector of equipment goods, their loss gets higher than 3 euros in the computer and social sectors. The loss is very large in the bank and insurance sector reaching 12.3 euros. In fact, the important squeezing in the sector of bank and insurance is not surprising as the white-collar population is very heterogenous in banks. Whereas most males are hired from the start as white collars, many females are hired in lower positions but have some carrier plans which allow them to become white collars after a while. In this context, good positions and job tenure allow males to get far higher wages than females within the white-collar category.

We also computed the test statistic to assess whether the model explains well the wage differences between males and females. Results show that the test statistic is below the 5% threshold (value 0.46136) for equipment goods, computers and social sector. Hence, the uniform discrimination and the glass ceiling effects would be enough to explain the wage differences in these sectors. By contrast, the test statistic is far above the 5% threshold for bank and insurance. This suggests that some other types of mechanisms are at stake in that sector. A reason may be that bank and insurance do not have the same union agreements.

We then tried to assess whether one type of discrimination only (uniform or glass ceiling) is enough to explain the difference in wage distributions between males and females for sectors for which the test statistic did not reject the specification when the two types of discrimination are included. When introducing only a translation, the specification is heavily rejected in all sectors (see Table 12). When introducing only the squeezing, the specification is also rejected although not so heavily in the equipment goods and social sectors (see Table 13). This suggests that both a uniform discrimination and a glass ceiling would occur. We finally tried to explain the difference in wage distributions with another specification, namely a uniform discrimination (translation) coupled with a discrimination that makes females loose a given share of their wage (dilatation). This is a transformation of values that writes $z_\beta = ay + b$, whereas there is no transformation of ranks. The test statistic reported in Table 14 accepts the specification only for the social sector although the statistic value is far larger than for the specification with translation and squeezing. This suggests that the glass ceiling is more compatible with the data than a discrimination yielding a wage proportional loss.

8 Conclusion

In this paper, we were interested in comparing the distributions of a continuous outcome between two groups. We considered specifications where the two distributions are related through some parametrized transformations of values and ranks. This type of specifications covers cases where the outcome is observed only for some selected individuals and the selection depends on the rank of the individuals in the distribution of their group. In fact, our specification can result from a specific class of theoretical models that we explicited.

We proposed a method based on quantiles to estimate the parameters of the value and rank transformations. This method can be decomposed in two steps. First, we estimated the quantiles of the two distribution non parametrically. Second, we minimized the distance between the transformed quantiles of the two distributions. We then derived the asymptotic distribution of the estimators and proposed a specification test that allows to assess whether the model specification is compatible with the data.

Monte Carlo simulations were conducted to evaluate the quality of our method. They showed that the estimators perform very well as long as the number of observations in each group is at least a few thousands. We finally applied our method to study the differences in the wage distributions between males and females. The goal was to quantify the effect of the glass ceiling and a uniform discrimination against females. Results showed that both types of discriminations lead to a wage loss for females.

Our method can be extended in several ways. We plan to include individual explanatory variables in the model. In that case, the quantiles functions of the two distributions will be conditional on these variables and may be estimated semi-parametrically using quantile regressions instead of sample quantile as we did in our setting. We also want to examine how the model can be written in a intertemporal framework where the distributions of the two groups are observed at several periods in time.

9 Appendix: asymptotics

We first propose two lemmas which will be used in the proofs proposed in the appendix.

Lemma 4 *Under assumptions A1-A5b,* $\widehat{\lambda}_j^{(k)}$ *converges uniformely in probability to* $\lambda_j^{(k)}$*, where* $\widehat{\lambda}_j^{(k)} = \frac{d^k \widehat{\lambda}_j}{du^k}$ *and* $\lambda_j^{(k)} = \frac{d^k \lambda_j}{du^k}$*, for* $k = 0, 1, 2$.

Proof. Denote $k_j(u,t) = \frac{\partial K_{N_j}}{\partial t}(u,t)$. We have:

$$\widehat{\lambda}_j^{(k)}(u) - \lambda_j^{(k)}(u) = S_j^k(u) + D_j^k(u)$$

with $S_j^k(u) = \int_0^1 \left[\widetilde{\lambda}_j(t) - \lambda_j(t)\right] \frac{\partial^k k_j}{\partial u^k}(u,t)\, dt$ the smoothing of the error between the sample quantile and the true quantile value, and $D_j^k(u) = \int_0^1 \lambda_j(t) \frac{\partial^k k_j}{\partial u^k}(u,t)\, dt - \lambda_j^{(k)}(u)$ a deterministic bias. We have:

$$\left|S_j^k(u)\right| \leqslant M^k(u, N_j) \int_0^1 \left|\widetilde{\lambda}_j(t) - \lambda_j(t)\right| dt$$

Corollary 21.5 p307 in Van der Vaart (1998) ensures that $\underset{u\in[0,1]}{sup} \left|\widetilde{\lambda}_j(u) - \lambda_j(u)\right| = o\left(N_j^{-\alpha}\right)$ for any $\alpha < 1/2$ as N_j tends to infinity. We can choose $\alpha = \theta_k + \varepsilon < 1/2$ with $\varepsilon > 0$ as $\theta_k < 1/2$. Hence, $\int_0^1 \left|\widetilde{\lambda}_j(t) - \lambda_j(t)\right| dt = o\left(N_j^{-\theta_k - \varepsilon}\right)$. Using A5b or A5c, we get $\underset{u\in[0,1]}{sup} \left|S_j^k\right| \leqslant o\left(N_j^{-\varepsilon}\right) \xrightarrow{P} 0$. Using A5a or A5d, we also have: $\underset{u\in[0,1]}{sup} \left|D_j^k\right| \xrightarrow{P} 0$. Hence:

$$\underset{u\in[0,1]}{sup} \left|\widehat{\lambda}_j^{(k)}(u) - \lambda_j^{(k)}(u)\right| \leqslant \underset{u\in[0,1]}{sup} \left|S_j^k(u)\right| + \underset{u\in[0,1]}{sup} \left|D_j^k(u)\right| \xrightarrow{P} 0$$

and $\widehat{\lambda}_j^{(k)} \xrightarrow{P} \lambda_j^{(k)}$ uniformely. ■

This result can be used to prove the following lemma:

Lemma 5 *For any sequence* $(\beta_n, \gamma_n) \in \Phi$*, where* n *is a bivariate index for the number of observations in the two groups* $n = (N_1, N_2)$ *with* $N_1 = 1, 2, ...$ *and* $N_2 = 1, 2, ...$*, such that* $(\beta_n, \gamma_n) \xrightarrow{P} (\beta_0, \gamma_0)$*, we have:*

$$\begin{array}{rcl}
\widehat{C}(\beta_n, \gamma_n) & \xrightarrow{P} & C(\beta_0, \gamma_0) = 0 \\
\left.\frac{\partial \widehat{C}}{\partial(\beta,\gamma)'}\right|_{(\beta_n,\gamma_n)} & \xrightarrow{P} & \left.\frac{\partial C}{\partial(\beta,\gamma)'}\right|_{(\beta_0,\gamma_0)} \\
\left.\frac{\partial^2 \widehat{C}}{\partial(\beta,\gamma)\partial(\beta,\gamma)'}\right|_{(\beta_n,\gamma_n)} & \xrightarrow{P} & \left.\frac{\partial^2 C}{\partial(\beta,\gamma)\partial(\beta,\gamma)'}\right|_{(\beta_0,\gamma_0)}
\end{array}$$

Proof. We now show that: $\widehat{C}_1(\beta_n,\gamma_n) \xrightarrow{P} C_1(\beta_0,\gamma_0)=0$ (the proof is similar for $\widehat{C}_2(\beta,\gamma)$). First note that we have: $\widehat{C}_1(\beta,\gamma)=\int_0^1 w_1(t)\, h(\beta,\gamma,t)^2\, dt$ with $h(\beta,\gamma,t)=z_{\beta,2}\left(\widehat{\lambda}_1(u_{\gamma,1}(t))\right)-\widehat{\lambda}_2(r_{\gamma,2}(u_{\gamma,1}(t)))$. Hence:

$$\begin{aligned}\widehat{C}_1(\beta_n,\gamma_n)-C_1(\beta_0,\gamma_0) &= \int_0^1 w_1(t)\left[h(\beta_n,\gamma_n,t)^2-h(\beta_0,\gamma_0,t)^2\right]dt \\ &= \int_0^1 w_1(t)\left[h(\beta_n,\gamma_n,t)+h(\beta_0,\gamma_0,t)\right]\left[h(\beta_n,\gamma_n,t)-h(\beta_0,\gamma_0,t)\right]dt\end{aligned}$$

As Φ is compact and and z is continous in all its arguments, $z_{\beta_n,2}$ is bounded. Moreover, $z_{\beta_0,2}$, $\widehat{\lambda}_1$ and λ_1 are also bounded. Thus there is an m_1 (which does not depend on n) such that $|h(\beta_n,\gamma_n,t)+h(\beta_0,\gamma_0,t)|\leqslant m_1$. Hence:

$$\widehat{C}_1(\beta_n,\gamma_n)\leq m_1\int_0^1 w_1(t)\,|h(\beta_n,\gamma_n,t)-h(\beta_0,\gamma_0,t)|\,dt$$

We have:

$$\begin{aligned}|h(\beta_n,\gamma_n,t)-h(\beta_0,\gamma_0,t)| \leq\ & \left|z_{\beta_n,2}\left(\widehat{\lambda}_1(u_{\gamma_n,1}(t))\right)-z_{\beta_n,2}\left(\lambda_1(u_{\gamma_n,1}(t))\right)\right| \\ &+\left|z_{\beta_n,2}\left(\lambda_1(u_{\gamma_n,1}(t))\right)-z_{\beta_n,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)\right| \\ &+\left|z_{\beta_n,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)-z_{\beta_0,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)\right|\end{aligned}$$

Using the mean value theorem and the boundedness of $\frac{\partial z_{\beta_n,2}}{\partial x}$, there is an m_2 such that:

$$\left|z_{\beta_n,2}\left(\widehat{\lambda}_1(u_{\gamma_n,1}(t))\right)-z_{\beta_n,2}\left(\lambda_1(u_{\gamma_n,1}(t))\right)\right|\leqslant m_2\left|\widehat{\lambda}_1(u_{\gamma_n,1}(t))-\lambda_1(u_{\gamma_n,1}(t))\right|$$

Using lemma 4, we have: $\underset{u\in[0,1]}{sup}\left|\widehat{\lambda}_1(u_{\gamma_n,1}(t))-\lambda_1(u_{\gamma_n,1}(t))\right|\xrightarrow{P}0$. Thus, we have:

$$\int_0^1 w_1(t)\left|z_{\beta_n,2}\left(\widehat{\lambda}_1(u_{\gamma_n,1}(t))\right)-z_{\beta_n,2}\left(\lambda_1(u_{\gamma_n,1}(t))\right)\right|dt\xrightarrow{P}0$$

As $\frac{d\lambda_1}{du}$ is bounded thanks to assumption A6, there is also an m_3 such that:

$$\left|z_{\beta_n,2}\left(\lambda_1(u_{\gamma_n,1}(t))\right)-z_{\beta_n,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)\right|\leqslant m_3\left|u_{\gamma_n,1}(t)-u_{\gamma_0,1}(t)\right|$$

Using the definition of $u_{\gamma,1}$ and the fact that $\gamma_n\xrightarrow{P}\gamma_0$, it is easy to show that:

$$\int_0^1 w_1(t)\left|u_{\gamma_n,1}(t)-u_{\gamma_0,1}(t)\right|dt\xrightarrow{P}0$$

and thus:

$$\int_0^1 w_1(t)\left|z_{\beta_n,2}\left(\lambda_1(u_{\gamma_n,1}(t))\right)-z_{\beta_n,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)\right|dt\xrightarrow{P}0$$

Finally, as $\frac{\partial z_{\beta,2}}{\partial\beta}$ is bounded, there is an m_4 such that:

$$\left|z_{\beta_n,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)-z_{\beta_0,2}\left(\lambda_1(u_{\gamma_0,1}(t))\right)\right|\leq m_4\,|\beta_n-\beta_0|$$

As $\beta_n\xrightarrow{P}\beta_0$, we have:

$$\int_0^1 w_1(t)\,|\beta_n-\beta_0|\,dt\xrightarrow{P}0$$

and thus:

$$\int_0^1 \left|z_{\beta_n,2}\left(\lambda_1\left(u_{\gamma_0,1}(t)\right)\right) - z_{\beta_0,2}\left(\lambda_1\left(u_{\gamma_0,1}(t)\right)\right)\right| dt \xrightarrow{P} 0$$

Combining all the intermediary results, we finally get:

$$\int_0^1 w_1(t)\left|z_{\beta_n,2}\left(\widehat{\lambda}_1\left(u_{\gamma_n,1}(t)\right)\right) - z_{\beta_0,2}\left(\lambda_1\left(u_{\gamma_0,1}(t)\right)\right)\right| dt \xrightarrow{P} 0$$

In the same way, we can write that:

$$\begin{aligned}
\widehat{\lambda}_2\left(r_{\gamma_n,2}\left(u_{\gamma_n,1}(t)\right)\right) - \lambda_2\left(r_{\gamma_0,2}\left(u_{\gamma_0,1}(t)\right)\right) &= \widehat{\lambda}_2\left(r_{\gamma_n,2}\left(u_{\gamma_n,1}(t)\right)\right) - \lambda_2\left(r_{\gamma_n,2}\left(u_{\gamma_n,1}(t)\right)\right) \\
&\quad + \lambda_2\left(r_{\gamma_n,2}\left(u_{\gamma_n,1}(t)\right)\right) - \lambda_2\left(r_{\gamma_n,2}\left(u_{\gamma_0,1}(t)\right)\right) \\
&\quad + \lambda_2\left(r_{\gamma_n,2}\left(u_{\gamma_0,1}(t)\right)\right) - \lambda_2\left(r_{\gamma_0,2}\left(u_{\gamma_0,1}(t)\right)\right)
\end{aligned}$$

Using the same line of proof as above using the properties of $r_{\gamma,2}$ instead of $z_{\beta,2}$, we get:

$$\int_0^1 w_1(t)\left|\widehat{\lambda}_2\left(r_{\gamma_n,2}\left(u_{\gamma_n,1}(t)\right)\right) - \lambda_2\left(r_{\gamma_0,2}\left(u_{\gamma_0,1}(t)\right)\right)\right| dt \xrightarrow{P} 0$$

This is enough to show that: $\widehat{C}_1(\beta_n,\gamma_n) \xrightarrow{P} 0$.

The same line of argument applies to show that: $\left.\frac{\partial \widehat{C}}{\partial(\beta,\gamma)'}\right|_{(\beta_n,\gamma_n)} \xrightarrow{P} \left.\frac{\partial C}{\partial(\beta,\gamma)'}\right|_{(\beta_0,\gamma_0)}$ and $\left.\frac{\partial^2 \widehat{C}}{\partial(\beta,\gamma)\partial(\beta,\gamma)'}\right|_{(\beta_n,\gamma_n)} \xrightarrow{P} \left.\frac{\partial^2 C}{\partial(\beta,\gamma)\partial(\beta,\gamma)'}\right|_{(\beta_0,\gamma_0)}$ using the boundedness of first, second and third-order derivatives of $z_{\gamma_j,2}$, $r_{\gamma_j,2}$ and λ_j. ■

9.1 Consistency of the estimated parameters (theorem 1)

We show the consistency of the estimated parameters using a reductio ad absurdum.
We suppose that $\left(\widehat{\beta},\widehat{\gamma}\right) \nrightarrow (\beta_0,\gamma_0)$. The identification assumption *A3* yields that: $\exists!\,(\beta_0,\gamma_0)\,|C(\beta_0,\gamma_0) = 0$. Since C is continuous and $C(\beta_0,\gamma_0) = 0$, then $C\left(\widehat{\beta},\widehat{\gamma}\right) \nrightarrow 0$. This means that: $\exists \varepsilon,\eta > 0$ such that $\forall N_0$, $\exists\ N > N_0|$ $P\left(C\left(\widehat{\beta},\widehat{\gamma}\right) > \eta\right) > \varepsilon$.

Using Lemma (4), we also have that: $\widehat{\lambda}_j \xrightarrow{P} \lambda_j$ uniformely in u for $j = 1,2$. This yields that for any (β,γ), $\widehat{C}(\beta,\gamma) \xrightarrow{P} C(\beta,\gamma)$. Suppose that the values of $(\beta,\gamma) \in \Phi$ where Φ is a compact set. Hence, $\widehat{C}(\beta,\gamma) \xrightarrow{P} C(\beta,\gamma)$ uniformely on Φ. Hence:

$$\exists N_1|\,\forall N > N_1,\ \forall(\beta,\gamma) \in \Phi,\ P\left(\left|\widehat{C}(\beta,\gamma) - C(\beta,\gamma)\right| > \frac{\eta}{2}\right) < \frac{\varepsilon}{2} \tag{21}$$

We fix $N_0 = N_1$ and select $N > N_0$ such that $P\left(C\left(\widehat{\beta},\widehat{\gamma}\right) > \eta\right) > \varepsilon$.
Applying (21) to $\left(\widehat{\beta},\widehat{\gamma}\right)$, we get: $P\left(\left|\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right) - C\left(\widehat{\beta},\widehat{\gamma}\right)\right| > \frac{\eta}{2}\right) < \frac{\varepsilon}{2}$.
Applying (21) to (β_0,γ_0), we get: $P\left(\left|\widehat{C}(\beta_0,\gamma_0)\right| > \frac{\eta}{2}\right) < \frac{\varepsilon}{2}$.
As $\widehat{\beta},\widehat{\gamma}$ minimizes $\widehat{C}$, we have $0 \leq \widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right) \leq \widehat{C}(\beta_0,\gamma_0)$ and we get: $P\left(\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right) > \frac{\eta}{2}\right) < \frac{\varepsilon}{2}$.

We have: $C\left(\widehat{\beta},\widehat{\gamma}\right)=C\left(\widehat{\beta},\widehat{\gamma}\right)-\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)+\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)$. Hence, $C\left(\widehat{\beta},\widehat{\gamma}\right)>\eta$, implies: $\left|C\left(\widehat{\beta},\widehat{\gamma}\right)-\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)\right|>\frac{\eta}{2}$ or $\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)>\frac{\eta}{2}$. This yields:

$$\begin{aligned} P\left(C\left(\widehat{\beta},\widehat{\gamma}\right)>\eta\right) &< P\left(\left\{\left|C\left(\widehat{\beta},\widehat{\gamma}\right)-\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)\right|>\frac{\eta}{2}\right\}\cup\left\{\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)>\frac{\eta}{2}\right\}\right) \\ &\leq P\left(\left|\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)-C\left(\widehat{\beta},\widehat{\gamma}\right)\right|>\frac{\eta}{2}\right)+P\left(\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right)>\frac{\eta}{2}\right)<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon \end{aligned}$$

This contradicts the fact that $P\left(C\left(\widehat{\beta},\widehat{\gamma}\right)>\eta\right)>\varepsilon$. ■

9.2 Asymptotic law of the estimated parameters (theorem 2)

In order to derive the asymptotic law of the estimated parameters, we do a Taylor expansion:

$$\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\widehat{\beta},\widehat{\gamma})}=\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\widehat{\beta},\widehat{\gamma})}-\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)}+\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)}$$

We have $\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\widehat{\beta},\widehat{\gamma})}=0$ since $\left(\widehat{\beta},\widehat{\gamma}\right)$ is the minimum of $\widehat{C}$. We do a Taylor expansion of the first right-hand side term and we get:

$$0=\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)}+\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}^T\left.\frac{\partial^2\widehat{C}}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\overline{\beta},\overline{\gamma})} \tag{22}$$

with $\overline{\beta}$ between β_0 and $\widehat{\beta}$, and $\overline{\gamma}$ between γ_0 and $\widehat{\gamma}$. Hence:

$$\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}=-\left(\left.\frac{\partial^2\widehat{C}}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\overline{\beta},\overline{\gamma})}\right)^{-1}\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)}$$

Using lemma 5, we get that:

$$\left(\left.\frac{\partial^2\widehat{C}}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)'}\right|_{(\overline{\beta},\overline{\gamma})}\right)^{-1}\xrightarrow{P}\left(\left.\frac{\partial^2 C}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)'}\right|_{(\beta_0,\gamma_0)}\right)^{-1}$$

We now establish the asymptotic law of $\left.\frac{\partial\widehat{C}}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)}=\left.\frac{\partial\widehat{C}_1}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)}+\left.\frac{\partial\widehat{C}_2}{\partial(\beta,\gamma)}\right|_{(\beta_0,\gamma_0)}$. Consider first the derivative of $\widehat{C}_1$with respect to γ:

$$\begin{aligned}\left.\frac{\partial\widehat{C}_1}{\partial\gamma}\right|_{(\beta_0,\gamma_0)} &= \int_0^1 w_1(t)\left.\frac{\partial\left[z_{\beta,2}\circ\underline{\widehat{\lambda}}_{\gamma,1}(t)-\widehat{\lambda}_2\circ\underline{r}_{\gamma,2}(t)\right]^2}{\partial\gamma}\right|_{(\beta_0,\gamma_0)}\mathrm{d}t \\ &= -2\int_0^1 w_1(t)\left(\begin{array}{c}\left.\frac{\partial r_{\gamma,2}}{\partial\gamma}\right|_{u_{\gamma_0,1}(t)}\widehat{\lambda}_2'\circ\underline{r}_{\gamma_0,2}(t)\\ +\left.\frac{\partial(u_{\gamma,1}(t))}{\partial\gamma}\right|_{\gamma_0}\frac{\left[\underline{\widehat{\lambda}}_{\gamma_0,1}'(t)*z_{\beta_0,2}'\circ\underline{\widehat{\lambda}}_{\gamma_0,1}(t)-\underline{r}_{\gamma_0,2}'(t)*\widehat{\lambda}_2'\circ\underline{r}_{\gamma_0,2}(t)\right]}{\overline{u}_{\gamma_0,1}-\underline{u}_{\gamma_0,1}}\end{array}\right)\left[z_{\beta_0,2}\circ\underline{\widehat{\lambda}}_{\gamma_0,1}(t)-\widehat{\lambda}_2\circ\underline{r}_{\gamma_0,2}(t)\right] \mathrm{c}\end{aligned}$$

As the quantiles $\widehat{\lambda}_j$ and their derivatives $\widehat{\lambda}_j'$ converge to their true value, and $\frac{\partial\left(u_{\gamma_0,1}(\cdot)\right)}{\partial t}$ is bounded on $[0,1]$, the second term in the paranthesis on the right-hand side tends to zero. Hence, we get:

$$\left.\frac{\partial \widehat{C}_1}{\partial \gamma}\right|_{(\beta_0,\gamma_0)} \sim -2\int_0^1 w_1(t)\left(\left.\frac{\partial r_{\gamma,2}}{\partial \gamma}\right|_{u_{\gamma_0,1}(t)} \widehat{\lambda}_2' \circ \underline{r}_{\gamma_0,2}(t)\right)\left[z_{\beta_0,2}\circ \underline{\widehat{\lambda}}_{\gamma_0,1}(t) - \widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t)\right] \mathrm{d}t$$

Deriving $\widehat{C}_1$ with respect to β, we get:

$$\left.\frac{\partial \widehat{C}_1}{\partial \beta}\right|_{(\beta_0,\gamma_0)} = 2\int_0^1 w_1(t) \left.\frac{\partial z_{\beta,2}}{\partial \beta}\right|_{\underline{\widehat{\lambda}}_{\gamma_0,1}(t)} \left[z_{\beta_0,2}\circ \underline{\widehat{\lambda}}_{\gamma_0,1}(t) - \widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t)\right] \mathrm{d}t$$

Let $\widehat{V}_{(\beta,\gamma),1}(t) = 2\begin{pmatrix} \left.\frac{\partial z_{\beta,2}}{\partial \beta}\right|_{\underline{\widehat{\lambda}}_{\gamma,1}(t)} \\ -\left.\frac{\partial r_{\gamma,2}}{\partial \gamma}\right|_{u_{\gamma,1}(t)} \widehat{\lambda}_2' \circ \underline{r}_{\gamma,2}(t)\end{pmatrix}$, we have

$$\left.\frac{\partial \widehat{C}_1}{\partial (\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} \sim \int_0^1 w_1(t)\, \widehat{V}_{(\beta_0,\gamma_0),1}(t)\left[z_{\beta_0,2}\circ \underline{\widehat{\lambda}}_{\gamma_0,1}(t) - \widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t)\right] \mathrm{d}t$$

Since $\underline{\widehat{\lambda}}_{\gamma,1}(\cdot) \xrightarrow{P} \underline{\lambda}_{\gamma,1}(\cdot)$ and $\widehat{\lambda}_2'(\cdot) \xrightarrow{P} \lambda_2'(\cdot)$ uniformly on $[0,1]$, $\widehat{V}_{(\beta,\gamma),1}(\cdot) \xrightarrow{P} V_{(\beta,\gamma),1}(\cdot)$ uniformly on $[0,1]$, with $V_{(\beta,\gamma),1}(t) = 2\begin{pmatrix} \left.\frac{\partial z_{\beta,2}}{\partial \beta}\right|_{\underline{\lambda}_{\gamma,1}(t)} \\ -\left.\frac{\partial r_{\gamma,2}}{\partial \gamma}\right|_{u_{\gamma,1}(t)} \lambda_2' \circ \underline{r}_{\gamma,2}(t)\end{pmatrix}$. Using also (2), we get:

$$\left.\frac{\partial \widehat{C}_1}{\partial (\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} \sim \int_0^1 w_1(t)\, V_{(\beta_0,\gamma_0),1}(t)\left[z_{\beta_0,2}\circ \underline{\widehat{\lambda}}_{\gamma_0,1}(t) - z_{\beta_0,2}\circ \underline{\lambda}_{\gamma_0,1}(t) - \left(\widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t) - \lambda_2 \circ \underline{r}_{\gamma_0,2}(t)\right)\right] \mathrm{d}t \tag{23}$$

We can then apply Theorem 1.3 p429 in Falk (1985) to the distribution $\underline{f}_{\gamma_0,1}$ for $t\in[0,1]$:

$$\sqrt{\widehat{N}_{\gamma_0,1}}\underline{f}_{\gamma_0,1}\left(\underline{\lambda}_{\gamma_0,1}(t)\right)\left(\underline{\widehat{\lambda}}_{\gamma_0,1}(t) - \underline{\lambda}_{\gamma_0,1}(t)\right) \overset{d}{\Longrightarrow} B_{11}(t) \tag{24}$$

where B_{11} is a Brownian bridge. Thus, using Slutsky's theorem,

$$\sqrt{\widehat{N}_{\gamma_0,1}}\left(z_{\beta_0,2}\circ \underline{\widehat{\lambda}}_{\gamma_0,1}(t) - z_{\beta_0,2}\circ \underline{\lambda}_{\gamma_0,1}(t)\right) \overset{d}{\Longrightarrow} \frac{z_{\beta_0,2}'\circ \underline{\lambda}_{\gamma_0,1}(t)}{\underline{f}_{\gamma_0,1}\circ \underline{\lambda}_{\gamma_0,1}(t)} B_{11}(t)$$

Consider the quantile function $\lambda_2 \circ \underline{r}_{\gamma_0,2}(t)$ for $t\in[0,1]$, the corresponding density is given by $f_2\left(\lambda_2\circ \underline{r}_{\gamma_0,2}(t)\right)/\underline{r}_{\gamma_0,2}'(t)$.

Applying Theorem 1.3 p429 in Falk (1985) to this distribution, we get:

$$\sqrt{\widehat{N}_{\gamma_0,2}}\frac{f_2\left(\lambda_2\circ \underline{r}_{\gamma_0,2}(t)\right)}{\underline{r}_{\gamma_0,2}'(t)}\left(\widehat{\lambda}_2\circ \underline{r}_{\gamma_0,2}(t) - \lambda_2\circ \underline{r}_{\gamma_0,2}(t)\right) \overset{d}{\Longrightarrow} B_{21}(t) \tag{25}$$

where B_{21} is a Brownian bridge independent of B_{11}. Thus,

$$\sqrt{\widehat{N}_{\gamma_0,2}}\left(\widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t) - \lambda_2 \circ \underline{r}_{\gamma_0,2}(t)\right) \overset{d}{\Longrightarrow} \frac{\underline{r}'_{\gamma_0,2}(t)}{f_2\left(\lambda_2 \circ \underline{r}_{\gamma_0,2}(t)\right)} B_{21}(t)$$

Define $\mu_1(t) = z_{\beta_0,2} \circ \underline{\lambda}_{\gamma_0,1}(t) = \lambda_2 \circ \underline{r}_{\gamma_0,2}(t)$, we have

$$\mu'_1(t) = \frac{z'_{\beta_0,2} \circ \underline{\lambda}_{\gamma_0,1}(t)}{\underline{f}_{\gamma_0,1} \circ \underline{\lambda}_{\gamma_0,1}(t)} = \frac{\underline{r}'_{\gamma_0,2}(t)}{f_2\left(\lambda_2 \circ \underline{r}_{\gamma_0,2}(t)\right)}$$

Hence

$$\sqrt{\widehat{N}_{\gamma_0,1}}\left(z_{\beta_0,2} \circ \widehat{\underline{\lambda}}_{\gamma_0,1}(t) - z_{\beta_0,2} \circ \underline{\lambda}_{\gamma_0,1}(t)\right) \overset{d}{\Longrightarrow} \mu'_1(t)\, B_{11}(t)$$
$$\sqrt{\widehat{N}_{\gamma_0,2}}\left(\widehat{\lambda}_2 \circ \underline{r}_{\gamma_0,2}(t) - \lambda_2 \circ \underline{r}_{\gamma_0,2}(t)\right) \overset{d}{\Longrightarrow} \mu'_1(t)\, B_{21}(t)$$

Hence, using (23), we obtain

$$\left.\frac{\partial \widehat{C}_1}{\partial (\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} \overset{d}{\Longrightarrow} \int_0^1 w_1(t)\, V_{(\beta_0,\gamma_0),1}(t)\, \mu'_1(t) \left[\frac{B_{11}(t)}{\sqrt{\widehat{N}_{\gamma_0,1}}} - \frac{B_{21}(t)}{\sqrt{\widehat{N}_{\gamma_0,2}}}\right] \mathrm{d}t$$

Denote:

$$B_1(t) = \sqrt{\frac{1}{\widehat{N}_{\gamma_0,1}} + \frac{1}{\widehat{N}_{\gamma_0,2}}} \left[\frac{B_{11}(t)}{\sqrt{\widehat{N}_{\gamma_0,1}}} - \frac{B_{21}(t)}{\sqrt{\widehat{N}_{\gamma_0,2}}}\right] \tag{26}$$

B_1 is a Brownian bridge. Hence,

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}} + \frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-\frac{1}{2}} \left.\frac{\partial \widehat{C}_1}{\partial (\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} \overset{d}{\Longrightarrow} \int_0^1 w_1(t)\, V_{(\beta_0,\gamma_0),1}(t)\, \mu'_1(t)\, B_1(t)\, \mathrm{d}t \tag{27}$$

Switching groups 1 and 2, it is possible to use the same line of proof to show that:

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}} + \frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-\frac{1}{2}} \left.\frac{\partial \widehat{C}_2}{\partial (\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} \overset{d}{\Longrightarrow} \int_0^1 w_2(t)\, V_{(\beta_0,\gamma_0),2}(t)\, \mu'_2(t)\, B_2(t)\, \mathrm{d}t \tag{28}$$

where $\mu_2(t) = z_{\beta_0,1} \circ \underline{\lambda}_{\gamma_0,2}(t)$, and B_2 is a brownian bridge defined by:

$$B_2(t) = \sqrt{\frac{1}{\widehat{N}_{\gamma_0,1}} + \frac{1}{\widehat{N}_{\gamma_0,2}}} \left[\frac{B_{22}(t)}{\sqrt{\widehat{N}_{\gamma_0,2}}} - \frac{B_{12}(t)}{\sqrt{\widehat{N}_{\gamma_0,1}}}\right] \tag{29}$$

with B_{12} and B_{12} are some independent Brownian bridges given by:

$$\sqrt{\widehat{N}_{\gamma_0,2}}\underline{f}_{\gamma_0,2}\left(\underline{\lambda}_{\gamma_0,2}(t)\right)\left(\widehat{\underline{\lambda}}_{\gamma_0,2}(t) - \underline{\lambda}_{\gamma_0,2}(t)\right) \overset{d}{\Longrightarrow} B_{22}(t) \tag{30}$$

$$\sqrt{\widehat{N}_{\gamma_0,1}} \frac{f_1\left(\lambda_1 \circ \underline{r}_{\gamma_0,1}(t)\right)}{\underline{r}'_{\gamma_0,1}(t)} \left(\widehat{\lambda}_1 \circ \underline{r}_{\gamma_0,1}(t) - \lambda_1 \circ \underline{r}_{\gamma_0,1}(t)\right) \overset{d}{\Longrightarrow} B_{12}(t) \tag{31}$$

We finally get:

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-\frac{1}{2}}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}$$

$$\overset{d}{\Longrightarrow}-\left(\left.\frac{\partial^2 C}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)}\right)^{-1}\left(\int_0^1 w_1(t)\,V_{(\beta_0,\gamma_0),1}(t)\,\mu_1'(t)\,B_1(t)\,\mathrm{d}t+\int_0^1 w_2(t)\,V_{(\beta_0,\gamma_0),2}(t)\,\mu_2'(t)\,B_2(t)\,\mathrm{d}t\right)$$

The estimated parameters asymptotically follow a normal law as they write as a sum of brownian bridges. In fact, we have:

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-\frac{1}{2}}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}\overset{d}{\Longrightarrow}N\left(0,\Gamma^{-1}\Omega\Gamma^{-1}\right)\tag{32}$$

with $\Gamma=\left.\frac{\partial^2 C}{\partial(\beta,\gamma)\partial(\beta,\gamma)'}\right|_{(\beta_0,\gamma_0)}$ and $\Omega=V\left(\int_0^1 w_1(t)\,V_{(\beta_0,\gamma_0),1}(t)\,\mu_1'(t)\,B_1(t)\,\mathrm{d}t+\int_0^1 w_2(t)\,V_{(\beta_0,\gamma_0),2}(t)\,\mu_2'(t)\,B_2(t)\,\mathrm{d}t\right)$.
We are now going to compute Γ and Ω.

We first calculate the expression of Γ. We have:

$$\begin{aligned}
\left.\frac{\partial^2 C_1}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} &= \int_0^1 w_1(t)\left.\frac{\partial\left[z_{\beta,2}\circ\underline{\lambda}_{\gamma,1}(t)-\lambda_2\circ\underline{r}_{\gamma,2}(t)\right]^2}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)}\mathrm{d}t\\
&= \int_0^1 w_1(t)\,V_{(\beta_0,\gamma_0),1}(t)\left.\frac{\partial\left[z_{\beta,2}\circ\underline{\lambda}_{\gamma,1}(t)-\lambda_2\circ\underline{r}_{\gamma,2}(t)\right]}{\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)}\mathrm{d}t\\
&= \frac{1}{2}\int_0^1 w_1(t)\,V_{(\beta_0,\gamma_0),1}(t)\,V_{(\beta_0,\gamma_0),1}^T(t)\,\mathrm{d}t\\
\left.\frac{\partial^2 C_2}{\partial(\beta,\gamma)\,\partial(\beta,\gamma)^T}\right|_{(\beta_0,\gamma_0)} &= \frac{1}{2}\int_0^1 w_2(t)\,V_{(\beta_0,\gamma_0),2}(t)\,V_{(\beta_0,\gamma_0),2}^T(t)\,\mathrm{d}t
\end{aligned}$$

Hence:

$$\Gamma=\frac{1}{2}\int_0^1\left[w_1(t)\,V_{(\beta_0,\gamma_0),1}(t)\,V_{(\beta_0,\gamma_0),1}^T(t)+w_2(t)\,V_{(\beta_0,\gamma_0),2}(t)\,V_{(\beta_0,\gamma_0),2}^T(t)\right]\mathrm{d}t$$

We now compute Ω. Note first that B_1 and B_2 are tied. Indeed we show below that for a given $t\in[0,1]$, it is possible to find an $s\in[0,1]$ such that $B_1(t)=B_2(s)$. This value of s verifies the equation:

$$u_{\gamma_0,1}(t)=r_{\gamma_0,1}\left(u_{\gamma_0,2}(s)\right)=\underline{r}_{\gamma_0,1}(s)\tag{33}$$

or equivalently:

$$u_{\gamma_0,2}(s)=r_{\gamma_0,2}\left(u_{\gamma_0,1}(t)\right)=\underline{r}_{\gamma_0,2}(t)\tag{34}$$

Indeed, if s verifies (33), we have: $\underline{\lambda}_{\gamma_0,1}(t)=\lambda_1\left(u_{\gamma_0,1}(t)\right)=\lambda_1\left(\underline{r}_{\gamma_0,1}(s)\right)$. According to the definitions of B_{11} and B_{12} given respectively by (24) and (31), this yields: $B_{11}(t)=B_{12}(s)$. Since s verifies (34), we also have: $\underline{\lambda}_{\gamma_0,2}(s)=\lambda_2\left(u_{\gamma_0,2}(s)\right)=\lambda_2\left(\underline{r}_{\gamma_0,2}(t)\right)$. According to the definitions of B_{21} and B_{22} given respectively by (25) and (30), this yields: $B_{21}(t) = B_{22}(s)$. Using the expressions of B_1 and B_2 given by (26) and (29), we finally obtain: $B_1(t) = B_2(s)$. Using (33), we thus have for all $t \in [0,1]$:

$$B_2(s) = B_1\left(\frac{\underline{r}_{\gamma_0,1}(s) - \underline{u}_{\gamma_0,1}}{\overline{u}_{\gamma_0,1} - \underline{u}_{\gamma_0,1}}\right) \tag{35}$$

We are now going to rewrite the integral $\int_0^1 w_2(s)\, V_{(\beta_0,\gamma_0),2}(s)\, \mu_2'(s)\, B_2(s)\, \mathrm{d}s$ as an integral of B_1 using (35). We make the change in variable: $t = \frac{\underline{r}_{\gamma_0,1}(s) - \underline{u}_{\gamma_0,1}}{\overline{u}_{\gamma_0,1} - \underline{u}_{\gamma_0,1}}$. We thus have:

$$\int_0^1 w_2(s)\, V_{(\beta_0,\gamma_0),2}(s)\, \mu_2'(s)\, B_2(s)\, \mathrm{d}s = \int_0^1 w_2(s)\, V_{(\beta_0,\gamma_0),2}(s)\, \mu_2'(s)\, \frac{ds}{dt} B_1(t)\, \mathrm{d}t \tag{36}$$

From (**??**) and (34), we also have:

$$\begin{aligned}
\mu_2(s) &= z_{\beta_0,1}\left[\underline{\lambda}_{\gamma_0,2}(s)\right] = z_{\beta_0,1}\left[\lambda_2\left(u_{\gamma_0,2}(s)\right)\right] \\
&= z_{\beta_0,1}\left[\lambda_2\left(\underline{r}_{\gamma_0,2}(t)\right)\right] = z_{\beta_0,1}\left[\mu_1(t)\right]
\end{aligned}$$

Deriving this expression with respect to t, we obtain:

$$\mu_2'(s)\frac{ds}{dt} = z_{\beta_0,1}'\left[\mu_1(t)\right] \mu_1'(t)$$

Hence:

$$\int_0^1 w_1(t)\, V_{(\beta_0,\gamma_0),1}(t)\, \mu_1'(t)\, B_1(t)\, \mathrm{d}t + \int_0^1 w_2(s)\, V_{(\beta_0,\gamma_0),2}(s)\, \mu_2'(s)\, B_2(s)\, \mathrm{d}s = \int_0^1 g(t)\, B_1(t)\, \mathrm{d}t \tag{37}$$

where:

$$\begin{aligned}
g(t) &= w_1(t)\, V_{(\beta_0,\gamma_0),1}(t)\, \mu_1'(t) + w_2(s(t))\, V_{(\beta_0,\gamma_0),2}(s(t))\, \mu_2'(s(t))\, \frac{ds(t)}{dt} \\
&= \left[w_1(t)\, V_{(\beta_0,\gamma_0),1}(t) + w_2(s(t))\, V_{(\beta_0,\gamma_0),2}(s(t))\, z_{\beta_0,1}'\left[\mu_1(t)\right]\right] \mu_1'(t)
\end{aligned}$$

with $s(t) = \frac{\underline{r}_{\gamma_0,2}(t) - \underline{u}_{\gamma_0,2}}{\overline{u}_{\gamma_0,2} - \underline{u}_{\gamma_0,2}}$. Let $T\{g\}(u) = \int_u^1 g(t)\, \mathrm{d}t - \int_0^1 t g(t)\, \mathrm{d}t$. We have:

$$\begin{aligned}
& \int_0^1 g(t)\, B_1(t)\, \mathrm{d}t \\
=& \int_0^1 g(t)\left(W_1(t) - tW_1(1)\right) \mathrm{d}t \\
=& \int_0^1 \int_0^t g(t)\, dW_1(u)\, \mathrm{d}t - \int_0^1 g(t)\, t\mathrm{d}t \int_0^1 \mathrm{d}W_1(u) \\
=& \int_0^1 T\{g\}(u)\, \mathrm{d}W_1(u)
\end{aligned}$$

where $W_1(\cdot)$ is a Wiener process. As for any u, $\mathrm{d}W_1(u)$ follows a normal law $N(0, du)$, and the processes $\mathrm{d}W_1(u)$, $u \in [0,1]$ are independent by definition, we finally get:

$$\Omega = V\left(\int_0^1 g(t)\, B_1(t)\, \mathrm{d}t\right) = \int_0^1 T\{g\}(u)\, T\{g\}(u)'\, \mathrm{d}u$$

9.3 Asymptotic law of the minimization criterium (theorem 3)

9.3.1 Theorem 3

We now establish the asymptotic law of the minimization criterium. Applying a Taylor expansion, we get:

$$\begin{aligned}\widehat{C}\left(\widehat{\beta},\widehat{\gamma}\right) &= \widehat{C}\left(\beta_0,\gamma_0\right)+\left.\frac{\partial \widehat{C}}{\partial\left(\beta,\gamma\right)^T}\right|_{\left(\beta_0,\gamma_0\right)}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}\\ &+\frac{1}{2}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}^T\left.\frac{\partial^2 \widehat{C}}{\partial\left(\beta,\gamma\right)\partial\left(\beta,\gamma\right)^T}\right|_{\left(\beta_0,\gamma_0\right)}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}+O\left[\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-3/2}\right]\end{aligned}\tag{38}$$

We first determine the asymptotic law of $\widehat{C}\left(\beta_0,\gamma_0\right)$. For that purpose we decompose it into $\widehat{C}_1\left(\beta_0,\gamma_0\right)$ and $\widehat{C}_2\left(\beta_0,\gamma_0\right)$. Injecting (2) into (11), we get:

$$\widehat{C}_1\left(\beta_0,\gamma_0\right)=\int_0^1 w_1\left(t\right)\left[z_{\beta_0,2}\left(\underline{\widehat{\lambda}}_{\gamma_0,1}\left(t\right)\right)-z_{\beta_0,2}\left(\underline{\lambda}_{\gamma_0,1}\left(t\right)\right)-\left[\widehat{\lambda}_2\left(\underline{r}_{\gamma_0,2}\left(t\right)\right)-\lambda_2\left(\underline{r}_{\gamma_0,2}\left(t\right)\right)\right]\right]^2\mathrm{d}t\tag{39}$$

Applying the same line of proof as when establishing (27), we obtain:

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1}\widehat{C}_1\left(\beta_0,\gamma_0\right)\overset{d}{\Longrightarrow}\int_0^1 w_1\left(t\right)\mu_1'\left(t\right)^2 B_1\left(t\right)^2\mathrm{d}t\tag{40}$$

Injecting (3) into (13), we get:

$$\widehat{C}_2\left(\beta_0,\gamma_0\right)=\int_0^1 w_2\left(t\right)\left[\widehat{\lambda}_1\left(\underline{r}_{\gamma_0,1}\left(t\right)\right)-\lambda_1\left(\underline{r}_{\gamma_0,1}\left(t\right)\right)-\left[z_{\beta_0,1}\left(\underline{\widehat{\lambda}}_{\gamma_0,2}\left(t\right)\right)-z_{\beta_0,1}\left(\underline{\lambda}_{\gamma_0,2}\left(t\right)\right)\right]\right]^2\mathrm{d}t\tag{41}$$

Applying the same line of proof as when establishing (28), we obtain:

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1}\widehat{C}_2\left(\beta_0,\gamma_0\right)\overset{d}{\Longrightarrow}\int_0^1 w_2\left(t\right)\mu_2'\left(t\right)^2 B_2\left(t\right)^2\mathrm{d}t\tag{42}$$

Combining (40) and (42), we get:

$$\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1}\widehat{C}\left(\beta_0,\gamma_0\right)\overset{d}{\Longrightarrow}\int_0^1 w_1\left(t\right)\mu_1'\left(t\right)^2 B_1\left(t\right)^2\mathrm{d}t+\int_0^1 w_2\left(t\right)\mu_2'\left(t\right)^2 B_2\left(t\right)^2\mathrm{d}t\tag{43}$$

Using the same line of argument as that to get from (36) to (37), we obtain:

$$\int_0^1 w_1\left(t\right)\mu_1'\left(t\right)^2 B_1\left(t\right)^2\mathrm{d}t+\int_0^1 w_2\left(s\right)\mu_2'\left(s\right)^2 B_2\left(s\right)^2\mathrm{d}s=\int_0^1 h\left(t\right)B_1^2\left(t\right)\mathrm{d}t$$

where:

$$h\left(t\right)=\left[w_1\left(t\right)+w_2\left(s\left(t\right)\right)z_{\beta_0,1}'\left[\mu_1\left(t\right)\right]^2/s'\left(t\right)\right]\mu_1'\left(t\right)^2$$

Using (22), the second right-hand side term in (38) rewrites:

$$-\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}^T\left.\frac{\partial^2 C}{\partial\left(\beta,\gamma\right)\partial\left(\beta,\gamma\right)^T}\right|_{\left(\overline{\beta},\overline{\gamma}\right)}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{\gamma}-\gamma_0\end{pmatrix}$$

It is equivalent to:

$$-\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}^T \Gamma \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}$$

The third right-hand side term in (38) is equivalent to:

$$\frac{1}{2}\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}^T \Gamma \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}$$

Hence, the sum of the second and third right-hand side terms in (38) is equivalent to:

$$-\frac{1}{2}\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}^T \Gamma \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}$$

We have: $\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-\frac{1}{2}} \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix} \overset{d}{\Longrightarrow} Y$ with Y following a normal distribution $N\left(0,\Gamma^{-1}\Omega\Gamma^{-1}\right)$. Denote Λ the Cholesky matrix such that: $\Gamma^{-1}\Omega\Gamma^{-1}=\Lambda\Lambda'$ and Y such that: $Y=\Lambda X$, where $V(X)=I$. Hence:

$$-\frac{1}{2}\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1} \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}^T \Gamma \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix} \sim -\frac{1}{2}X'\Lambda'\Gamma\Lambda X$$

Consider an orthonormal decomposition of $\Lambda'\Gamma\Lambda$ such that $\Lambda'\Gamma\Lambda=\Xi D\Xi'$ where Ξ verifies $\Xi\Xi'=I$ and D is the diagonal matrix of the eigenvalues of $\Lambda'\Gamma\Lambda$. We have: $X'\Lambda'\Gamma\Lambda X=\widetilde{X}'D\widetilde{X}$ with $\widetilde{X}=\Xi X$. Here, $V\left(\widetilde{X}\right)=\Xi V\left(X\right)\Xi'=\Xi I\Xi'=\Xi\Xi'=I$. Thus, $\widetilde{X}'D\widetilde{X}=\sum\limits_k D_k\widetilde{X}_k^2$ where D_k is the k^{th} eigenvalue of $\Lambda'\Gamma\Lambda$ (this is a special case of the Cochran's theorem). Finally:

$$-\frac{1}{2}\left(\frac{1}{\widehat{N}_{\gamma_0,1}}+\frac{1}{\widehat{N}_{\gamma_0,2}}\right)^{-1} \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix}^T \Gamma \begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\gamma}-\gamma_0 \end{pmatrix} \overset{d}{\Longrightarrow} -\frac{1}{2}\sum_k D_k\widetilde{X}_k^2$$

where $-\frac{1}{2}\sum\limits_k D_k\widetilde{X}_k^2$ is a weighted sum of chi-squares.

References

[1] Abadie, Angrist and Imbens (2002), "Intrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings", *Econometrica*, 70(1), pp. 91-117.

[2] Albrecht J., Björklund A. and S. Vroman (2003), "Is There a Glass Ceiling in Sweden?", *Journal of Labour Economics*, 21(1), pp. 145-177.

[3] Athey S. and G. Imbens (2006), "Identification and Inference in Nonlinear Differences-in-Differences Models", *Econometrica*, 74(2), pp. 431-497.

[4] Bloch D. A. and J.L. Garwirth (1968), "On a Simple Esytimate of the Reciprocal of the Density Function", *The Annals of Mathematical Statistics*, 39(3), pp. 1083-1085.

[5] Cheng C. and E. Parzen (1997), "Unified estimators of smooth quantile and quantile density functions", *Journal of Statistical Planning and Inference*, 59, pp. 291-307.

[6] Chernozhukov and Hansen (2005), "A IV model of Quantile Treatment Effects", *Econometrica*, 73(1), pp. 245-261.

[7] Combes PPh., Duranton G., Gobillon L., Puga D. and S. Roux (2008), "The productivity advantages of large cities: Distinguishing agglomeration from firm selection", mimeo.

[8] Doksum K. (1974), "Empirical Probability Plots ans Statistical Inference for Nonlinear Models in the Two-Sample Case", *The Annals of Statistics*, 2(2), p. 267-277.

[9] Duranton G. and D. Puga (2004), "Microfoundations of urban agglomeration economies", *Handbook of Regional and Urban Economics*, Volume 4, pp. 2063-2117, J.V. Henderson and J.F. Thisse (eds).

[10] Falk M. (1985), "Asymptotic Normality of the Kernel Quantile Estimator", *The Annals of Statistics*, 13(1), pp. 428-433.

[11] Firpo (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects", *Econometrica*, 75(1), pp. 259-276.

[12] Härdle (1990), *Applied Nonparametric Regressions*, 349p, Cambride University Press.

[13] Köenker (2005), *Quantile Regressions*, 366p, Cambride University Press.

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

[15] Knott M. (1974), "The Distribution of the Cramer-von Mises Statistic for Small Sample Sizes", *Journal of the Royal Statistical Society, Series B (Methodological)*, 36(3), pp. 430-438.

[16] Melitz M. (2003), "The Impact of Trade and Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, 71, pp. 1695-1725.

[17] Melitz M. and G. Ottaviano (2007), "Market Size, Trade, and Productivity", *Review of Economic Studies*, forthcoming.

[18] Siddiqui (1960), M.M., "Distribution of quantiles in samples from a bivariate population", *Journal of Research of the National Standards-B Mathematics and Mathematical Physics, 64b(3),* pp. 145-150.

[19] Syverson C. (2004), "Market Structure and Productivity: A Concrete Example", *The Journal of Political Economy*, 112(6), pp. 1181-1222.

[20] Van der Vaart A.W. (1998), *Asymptotic Statistics*, 443p, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.