DISCUSSION PAPER SERIES

IZA DP No. 8256

Extremal Quantile Regressions for Selection Models and the Black-White Wage Gap

Xavier D'Haultfœuille
Arnaud Maurel
Yichong Zhang

June 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Extremal Quantile Regressions for Selection Models and the Black-White Wage Gap

Xavier D'Haultfœuille
CREST

Arnaud Maurel
Duke University, NBER and IZA

Yichong Zhang
Duke University

Discussion Paper No. 8256
June 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 8256
June 2014

ABSTRACT

Extremal Quantile Regressions for Selection Models and the Black-White Wage Gap*

We consider the estimation of a semiparametric location-scale model subject to endogenous selection, in the absence of an instrument or a large support regressor. Identification relies on the independence between the covariates and selection, for arbitrarily large values of the outcome. In this context, we propose a simple estimator, which combines extremal quantile regressions with minimum distance. We establish the asymptotic normality of this estimator by extending previous results on extremal quantile regressions to allow for selection. Finally, we apply our method to estimate the black-white wage gap among males from the NLSY79 and NLSY97. We find that premarket factors such as AFQT and family background characteristics play a key role in explaining the level and evolution of the black-white wage gap.

JEL Classification: C21, C24, J31

Keywords: sample selection models, extremal quantile regressions, black-white wage gap

Corresponding author:

Arnaud Maurel
Duke University
Department of Economics
213 Social Sciences
Durham, NC 27708-0097
USA
E-mail: apm16@duke.edu

* We are grateful to Derek Neal for useful suggestions and providing us with the sample used in his 1996 *JPE* article. We also thank Peter Arcidiacono, Victor Chernozhukov, Ivan Fernandez-Val, Shakeeb Khan, Pat Kline, Adam Rosen, Lowell Taylor, Ed Vytlacil, and participants at various seminars and conferences for useful comments and suggestions.

1 Introduction

Endogenous selection has been recognized as one of the key methodological issues arising in the analysis of microeconomic data since the seminal articles of Gronau (1974) and Heckman (1974). The most common strategy to deal with selection is to rely on instruments that determine selection but not the potential outcome (see, among many others, Heckman, 1974, 1979, 1990, Ahn & Powell, 1993, Das *et al.* , 2003, and Vella, 1998 for a survey). However, in practice, valid instruments are generally difficult to find. Identification at infinity has been proposed in the literature as an alternative solution to the endogenous selection problem, in situations where one is primarily interested in estimating the effects of some covariates on a potential outcome. In particular, Chamberlain (1986) showed that if some individuals face an arbitrarily large probability of selection and the outcome equation is linear, then one can use these individuals to identify the effects of the covariates on the outcome of interest. Lewbel (2007) generalized this result by proving that identification can be achieved in the context of moment equality models, provided that a special regressor has a support which includes that of the error term from the selection equation (see Lewbel, 2014, for an overview of the special regressor method). Again, in many applications, such a regressor is hard to come by. In a recent article, D'Haultfoeuille & Maurel (2013a) have shown that identification in the absence of an instrument is in fact possible without such a covariate. The starting intuition is that, if selection is truly endogenous, then one can expect the effect of the outcome on selection to dominate those of the covariates for sufficiently large values of the outcome. Following this idea, they proved identification under the key condition that selection becomes independent of the covariates at the limit, i.e., when the *outcome* tends to the upper bound of its support.

This paper builds on this insight and develops a novel inference method for a class of semiparametric location-scale models subject to endogenous selection. Unlike prior estimation methods for sample selection models, we propose a distribution-free estimator that does not require an instrument for selection or a large support regressor. While D'Haultfoeuille & Maurel (2013a) prove identification for a nonparametric location-scale outcome equation, we rely instead on a semiparametric specification in order to obtain faster convergence rates. This is crucial for the practical usefulness of our method. Importantly, the model is left unrestricted otherwise. In particular, we do not restrict the selection process, apart from the independence at the limit condition mentioned above. In this paper, we interpret this condition in the context of standard selection models, and show that it translates into a restriction on the copula between the error terms of the outcome and selection equation. This restriction is mild provided that selection is endogenous, and holds in particular for all Gaussian copulas

with positive dependence. In this context, we show that linear quantile regressions, for large values of the quantile indices, allow us to recover some linear combinations of the covariates effects on the location and scale of the outcome. Those parameters can then be estimated in a second step by a simple minimum distance estimator, which combines the previous estimators for a range of quantile indices. This insight is important for at least two reasons. First, our estimator is simple to implement. In particular, unlike most of the existing semiparametric estimators for sample selection models, our estimator is not based on a nonparametric first step. Second, the asymptotic properties of extremal quantile regressions, that is quantile regressions applied to the tails, have been thoroughly studied in the case without selection in an important paper by Chernozhukov (2005).[1] This provides a very natural starting point to develop asymptotic inference in our setting. The estimators of the location and scale parameters can then be used to construct bounds on the quantile effects. We characterize the sharp bounds and further derive simpler outer bounds on which one can conduct inference using the methodology developed by Chernozhukov *et al.* (2013). It is worth noting that, while we use quantile regressions as a tool to circumvent the selection issue, we assume a linear location-scale specification for the potential outcome. This is dfferent in spirit from the methods proposed in the literature to estimate quantile regression models in the presence of sample selection (see notably Buchinsky, 1998, and Arellano & Bonhomme, 2011).

The main difficulty in establishing the asymptotic properties of our estimator is that because of selection, extremal conditional quantiles are not exactly linear here, but only equivalent to a linear form as the quantile index τ_n tends to zero. Hence, we face a bias-variance trade-off that is typical in non- or semiparametric analysis. Choosing a moderately small quantile index decreases the variance of the estimator, but this comes at the price of a higher bias. Conversely, choosing a very small quantile index mitigates the bias, but increases the variance. In the paper, we provide sufficient conditions under which both bias and variance vanish asymptotically, resulting in asymptotically normal and unbiased estimators. As in the case without selection examined by Chernozhukov (2005), the convergence rates are not standard, and depend on the tail behavior of the error term from the outcome equation. This is broadly similar to the convergence rates discussed in Andrews & Schafgans (1998), Schafgans & Zinde-Walsh (2002) and Khan & Tamer (2010), the main difference being that

[1] Formally, denoting by n the sample size and τ_n the quantile index, extremal quantile regressions correspond to τ_n-quantile regressions where τ_n tends to zero as the sample size n grows to infinity. In this paper, we focus on the intermediate order case, which corresponds to situations where $\tau_n \times n$ tends to infinity. See Chernozhukov & Du (2008) for a review of extremal quantile regressions. See also related work by Altonji *et al.* (2008), who derive the asymptotic properties of a nonparametric extremal quantile regression estimator. While their framework is very general, it cannot be readily extended to the case where the outcome is subject to sample selection.

the tail behavior of the outcome is going to play a key role here, rather than that of the covariates. Importantly, though, our asymptotic results suggest a rate-adaptive approach for inference, as in Khan & Tamer (2010) and Chernozhukov & Fernandez-Val (2011).

Asymptotic normality and unbiasedness of our estimators requires an appropriate choice of the quantile index, similarly to nonparametric kernel regressions that require an appropriate bandwidth choice. But contrary to the latter case, admissible rates of convergence towards zero for the quantile index depend in a complicated way on the data-generating process. An analogous issue arises in the estimation at infinity of the intercept of sample selection models (see Andrews & Schafgans, 1998 and Schafgans & Zinde-Walsh, 2002), as well as in the estimation of extreme value indices (see Drees & Kaufmann, 1998 and Danielsson *et al.* , 2001). This is a difficult problem. In the paper, we propose a heuristic data-driven procedure that selects the quantile index minimizing a criterion function capturing the trade-off between bias and variance. In particular, we use subsampling combined with a minimum distance estimator to proxy the bias term, which, in this setting, cannot be simply estimated. Monte Carlo simulation results show that the sampling distributions of our estimators are fairly well approximated by the asymptotic normal distributions, suggesting that our procedure performs well in practice.

Finally, we apply our method to the estimation of the black-white wage gap among males from the 1979 and 1997 cohorts of the National Longitudinal Survey of Youth (NLSY79 and NLSY97). Following Neal & Johnson (1996), we focus on the residual portion of the wage gap that remains after controlling for premarket factors. To the extent that black males are more likely to dropout from the labor market than white males, as was first pointed out in the influential work of Butler & Heckman (1977), correcting for selection is crucial for consistently estimating the black-white differential in terms of potential wages. Besides, evidence that the black-white employment gap has substantially widened over time (see, e.g., Juhn, 2003 and Neal & Rick, 2014) stresses the importance of dealing with selection in order to draw valid conclusions regarding the across-cohort evolution of the black-white wage gap. In this context, finding a valid instrument that affects selection but not potential wages is particularly challenging, making it desirable to use an estimation method that does not require such an instrument. For the NLSY79 cohort, we find a smaller residual wage gap (10.1%) than the one obtained using the imputation method of Neal & Johnson (1996) and Johnson *et al.* (2000), which is consistent with our approach being based on a weaker identifying restriction.[2] Overall, our estimates strengthen the key takeaway of Neal & Johnson (1996) by

[2]Other noteworthy papers analyzing the black-white wage gap while using imputation methods to correct for selection into the workforce include Brown (1984), Smith & Welch (1989), Juhn (2003), Neal (2004), Neal

providing evidence of an even more important role played by the black-white AFQT gap.

Turning to the evolution across the 1979 and 1997 cohorts, we find that there has been a slow convergence in the raw male black-white wage gap between 1990 and 2007 (-4.6 pp), and an even slower convergence in the residual portion of the wage gap that remains after controlling for premarket factors such as AFQT and family background (-1.2 pp). Interestingly, this provides evidence that premarket skills are a key component of the level as well as the evolution of the black-white wage gap. Besides, the fact that the wage gap which remains after accounting for differences in premarket factors is essentially stable after almost 20 years suggests that this residual portion of the wage gap is an important factor behind the slow convergence of the wages of blacks and whites.

The remainder of the paper is organized as follows. Section 2 presents the set-up and discusses the identification results. Section 3 defines the estimators and establishes the main asymptotic normality results. Section 4 discusses some Monte Carlo simulation results. Section 5 applies our method to the estimation of the black-white wage gap among males. Finally, Section 6 concludes. Additional details on the estimation procedure and the data, along with the proofs, are collected in the appendix.

2 The set-up and identification

2.1 Model and main result

Before presenting the model, let us introduce some notations and definitions. For any random variable U, we denote by F_U and S_U its cumulative distribution function (cdf.) and survival function, while Q_U denotes its quantile function, $Q_U(u) = \inf\{u : F_U(u) \geq \tau\}$. For more general increasing functions G, we let $G^{\leftarrow}(u) = \inf\{v : G(v) \geq u\}$, with the convention that $\inf \emptyset = +\infty$, denote its generalized inverse. Finally, we use in the following some notions from extreme value theory. A function F is regularly varying at $x \in \{0, +\infty\}$ with index $\alpha \in [-\infty, +\infty]$, and we write $F \in RV_\alpha(x)$, if for any $t > 0$, $\lim_{u \to x} F(tu)/F(u) = t^\alpha$, with the understanding that $t^\infty = \infty$ if $t > 1$ and $= 0$ if $1 > t > 0$ (and similarly for $\alpha = -\infty$). F is slowly varying at x if $F \in RV_0(x)$. We also say that a given cdf. F belongs to the domain of attraction of generalized extreme value distributions if there exists sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ and a cdf. G such that for any independent draws $(U_1, ..., U_n)$ from F, $b_n^{-1}(\max(U_1, ..., U_n) - a_n)$ converges in distribution to G. In such a case, G belongs to the family of generalized extreme value distributions.

(2006) and Neal & Rick (2014).

Let Y^* denote the outcome of interest and $X \in \mathbb{R}^d$ denote a vector of covariates, excluding the constant. We suppose that Y^* and X are related through the location-scale model

$$Y^* = X'\beta + (1 + X'\delta)\varepsilon, \tag{2.1}$$

where we suppose, without loss of generality, that $1 + X'\delta > 0$. Because we do not standardize ε, we can always suppose that there is no intercept and fix the constant in the multiplier of ε to one. Our focus throughout the paper is on the parameters β, δ, along with the quantile effects of X_j on Y^*. These quantile effects correspond to the effect on Y^* of an exogenous, infinitesimal, change of X_j, or a change from $X_j = 0$ to $X_j = 1$ if X_j is binary, for individuals at a given conditional quantile of Y^*. We do not seek to recover the distribution of ε, and in particular the intercept $E(\varepsilon)$, though this distribution will play an important role in our analysis.

We face a sample selection issue here as we only observe $(D, Y = DY^*, X)$, where D denotes the selection dummy. Importantly, we do not assume to have access to an instrument affecting D but not Y^*, nor do we require one of the covariates to have a large support. Instead, identification is achieved under the following conditions.

Assumption 1. *(Exogeneity)* $X \perp\!\!\!\perp \varepsilon$.

Assumption 2. *(Covariates)* X *has a compact support* $Supp(X)$. *Let* $\overline{X} = [1, X']'$, $\mathcal{Q}_X = E(\overline{X}\overline{X}')$ *is full rank.*

Assumption 3. *(Tail and regularity of the residual) (i)* $\sup(Supp(\varepsilon)) = \infty$, *(ii)* $S_{\exp(\varepsilon)}$ *is not slowly varying at infinity, (iii)* S_ε *is in the domain of attraction of generalized extreme value distributions and (iv) the distribution of* (X, ε) *conditional on* $D = 1$ *is dominated by a product measure. We denote by* $f_{\varepsilon|D=1,X}$ *and* $f_{Y|D=1,X}$ *the corresponding conditional densities.*

Assumption 4. *(Independence at infinity) There exists* $h > 0$ *such that for all* $x \in Supp(X)$,

$$\lim_{y \to \infty} P(D = 1 | X = x, Y^* = y) = h.$$

Assumption 1 is restrictive but commonly made in the context of selection models.[3] It is also weaker than the exogeneity assumption imposed, for instance, by Chamberlain (1986) or Ahn

[3]Notable exceptions include Das *et al.* (2003) and Lewbel (2007), who allow for endogenous regressors. However, the estimators proposed in these papers require an instrument for selection or a special regressor, respectively.

& Powell (1993), since we allow for heteroskedasticity here. The compact support condition in Assumption 2 is not required for identification but will be needed when using extremal quantile regression techniques. We maintain this assumption here for simplicity. Assumption 3-(ii) is satisfied if, for instance, $E(\exp(b\varepsilon)) < \infty$ for some $b > 0$. Note that this tail condition is fairly mild. For example, in the context of a wage equation where Y^* corresponds to the logarithm of the wage w, it is satisfied as long as $E(w^b) < \infty$ for a given $b > 0$. It follows that this condition holds even if wages exhibit very fat tails, for instance Pareto-like. Conditions (i) and (iii) are not necessary for identification but will be used subsequently. Condition (iii) is mild and satisfied by most of the standard continuous cdf., including the normal one. Condition (iv), which is very mild, is not needed for the identification of β and δ. It is only required to define the bounds on the quantile effects. Finally, Assumption 4 is our key identifying condition. We require selection to become independent of the covariates at infinity, that is conditional on having arbitrarily large outcomes. The underlying intuition is that, if selection is indeed endogenous, then one can expect the effect of the outcome on selection to dominate those of the covariates for sufficiently large values of the outcome. We come back to this condition in the following section, by discussing in detail several examples where this condition holds.

The first part of Theorem 2.1 below follows from Theorem 2.1 in D'Haultfoeuille & Maurel (2013a). They proved, in a nonparametric framework where $X'\beta$ and $X'\delta$ are replaced by two unrestricted functions $\psi(X)$ and $\sigma(X)$, that these functions are identified. Assumption 2 then ensures that β and δ are identified. The second part of Theorem 2.1 provides the sharp bounds on the quantile effects, which were not considered in D'Haultfoeuille & Maurel (2013a).

Theorem 2.1. *Under Assumptions 1-4, β and δ are identified. Moreover, the quantile effect $\Delta_{j\tau} = \partial Q_{Y^*|X}/\partial X_j(\tau|X)$[4] satisfies $\Delta_{j\tau} \in [\underline{\Delta}_{j\tau}, \overline{\Delta}_{j\tau}]$, with*

$$\begin{aligned}\underline{\Delta}_{j\tau} &= \beta_j + \delta_j \underline{F}_{\varepsilon}^{\leftarrow}\left[\tau - \mathbb{1}\{\delta_j > 0\}(1 - \underline{F}_{\varepsilon}(+\infty))\right], \\ \overline{\Delta}_{j\tau} &= \beta_j + \delta_j \underline{F}_{\varepsilon}^{\leftarrow}\left[\tau - \mathbb{1}\{\delta_j < 0\}(1 - \underline{F}_{\varepsilon}(+\infty))\right],\end{aligned}$$

and where

$$\underline{F}_{\varepsilon}(v) = \int_{-\infty}^{v} \sup_{x \in Supp(X)} \left[P(D=1|X=x)(1+x'\delta) f_{Y|D=1,X}(x'\beta + (1+x'\delta)u|x)\right] du.$$

[4]When X_j is binary, $\Delta_{j\tau}$ should rather be defined by $\Delta_{j\tau} = Q_{Y^*|X_{-j},X_j=1} - Q_{Y^*|X_{-j},X_j=0}$, where X_{-j} denotes all components of X except X_j. Under Assumption 1, both definitions coincide and are equal to $\beta_j + \delta_j Q_{\varepsilon}(\tau)$.

Finally, suppose that there exists $K > 0$ such that for all $(x, x') \in Supp(X)^2$ and all y large enough,

$$|P(D = 1|X = x, Y^* = y) - P(D = 1|X = x', Y^* = y)| \leq K\|x - x'\|, \tag{2.2}$$

where $\|x\|$ denotes the euclidian norm of x. Then the bounds $\underline{\Delta}_{j\tau}$ and $\overline{\Delta}_{j\tau}$ are sharp.

The underlying intuition of the identification result for (β, δ) is that under Assumption 4, the right tail of Y and Y^* are equivalent up to a multiplicative constant. It follows that we can use the conditional survival function of Y given X to uniquely recover the location and scale parameter, provided the tail of the residual is not too fat (Assumption 3-(ii)). Point identification of β and δ does not necessarily entail point identification of the quantile effects $\Delta_{j\tau}$, because of potential heteroskedasticity. Specifically, $\Delta_{j\tau}$ is point identified under partial homoskedasticity ($\delta_j = 0$), but only partially identified otherwise. This is due to the fact that $\Delta_{j\tau} = \beta_j + \delta_j Q_\varepsilon(\tau)$. Because of the missing data issue, the quantile $Q_\varepsilon(\tau)$ cannot be point identified in general.

An important limitation of the sharp bounds is that, to the best of our knowledge, no existing method can be readily used to conduct inference on them. Kitagawa (2010) provides some useful results, but in a simpler framework where X is discrete and without the need to estimate β and δ in a first step. On the other hand, we derive in Appendix C.2 the following outer bounds on $\Delta_{j\tau}$:

$$\begin{array}{rcl} \underline{\Delta}^o_{j\tau} & = & \beta_j + \delta_j \left[\mathbb{1}\{\delta_j > 0\}\underline{Q}^o_\varepsilon(\tau) + \mathbb{1}\{\delta_j < 0\}\overline{Q}^o_\varepsilon(\tau)\right] \\ \overline{\Delta}^o_{j\tau} & = & \beta_j + \delta_j \left[\mathbb{1}\{\delta_j > 0\}\overline{Q}^o_\varepsilon(\tau) + \mathbb{1}\{\delta_j < 0\}\underline{Q}^o_\varepsilon(\tau)\right] \end{array}, \tag{2.3}$$

with

$$\underline{Q}^o_\varepsilon(\tau) = \sup_{x \in \text{Supp}(X)} \frac{Q_{Y|D=1,X=x}\left(\frac{\tau - P(D=0|X=x)}{P(D=1|X=x)}\right) - x'\beta}{1 + x'\delta},$$

$$\overline{Q}^o_\varepsilon(\tau) = \inf_{x \in \text{Supp}(X)} \frac{Q_{Y|D=1,X=x}\left(\frac{\tau}{P(D=1|X=x)}\right) - x'\beta}{1 + x'\delta}.$$

The outer bounds take a more convenient form for inference than the sharp bounds. Besides, they actually coincide with these sharp bounds when the family of functions $e \mapsto P(D = 1|X = x)(1 + x'\delta) f_{Y|D=1,X}(x'\beta + (1 + x'\delta)e|x)$ indexed by x do not intersect.

2.2 The independence at infinity condition

Point identification of β and δ relies mostly on the independence at infinity assumption. To get a better sense of this condition, we discuss it below in the context of two common selection models. The first one is a threshold crossing model described in Assumption 5.

Assumption 5. *(i)* $D = \mathbb{1}\{\phi(X) - \eta \geq 0\}$ *with* $(\varepsilon, \eta) \perp\!\!\!\perp X$*, (ii)* $\inf_{x \in Supp(X)} F_\eta(\phi(x)) = \underline{v} > 0$*, (iii)* F_ε *and* F_η *are continuous and strictly increasing and the copula* C *of* $(-\varepsilon, \eta)$ *is differentiable with respect to its first argument.*

The first condition defines the selection model as a standard threshold crossing model. Importantly however, we do not add any instrument in this selection equation. The second condition ensures that $x \mapsto P(D = 1|X = x)$ is bounded below by a positive number. Note that this condition will typically hold if none of the covariates has a large support, which is precisely the type of situation we are interested in. In this context, Proposition 2.1 provides a restriction on C ensuring that Assumption 4 is satisfied. Hereafter, let $f_C(\tau) = \sup_{u \leq \tau, v \in [\underline{v}, 1]} |\partial_1 C(u, v) - 1|$, where $\partial_1 C$ denotes the partial derivative of C with respect to its first argument.

Proposition 2.1. *Suppose that Assumptions 2, 3 and 5 hold, and*

$$\lim_{\tau \to 0} f_C(\tau) = 0. \tag{2.4}$$

Then Assumption 4 is satisfied, and therefore β *and* δ *are identified.*

The key idea is that selection becomes independent of the covariates for large values of the outcome if selection is endogenous enough, in the sense that $(-\varepsilon, \eta)$ satisfies (2.4). To understand this condition better, it is useful to consider two extreme cases. In the perfect dependence case such that $\eta = -\varepsilon$, then $\partial_1 C(u, v) = 1$ for all $u < v$, so that (2.4) actually holds exactly for small values of τ. On the other hand, when η and $-\varepsilon$ are independent, $\partial_1 C(u, v) = v$, and $f_C(\tau) = 1 - \underline{v}$, which is positive except in the degenerate case where $D = 1$ almost surely. In between these two extreme cases, Table 1 provides examples of copulas that satisfy this constraint. Importantly, it holds for all Gaussian copulas with positive dependence.[5] It also holds for Archimedean copulas under a restriction on the behavior of the generator Ψ around 0. This restriction holds for instance for the Clayton copula, for which $\Psi(u) = (u^{-\theta} - 1)/\theta$, provided that $\theta > 0$. The Gumbel family is another popular Archimedean family of copulas that does not satisfy the restriction on Ψ, since Ψ is slowly varying at 0. However, Condition (2.4) still holds for some parameters of this family.

[5] Note that what is important here is the strength, but not the sign, of the dependence between η and $-\varepsilon$. The case of negative dependence could be addressed by replacing D by $1 - D$ (see D'Haultfoeuille & Maurel, 2013a for a discussion on this point).

Copula family	Restriction ensuring (2.4)
Gaussian $C(u,v;\rho)$	$\rho > 0$
Archimedean $C(u,v;\Psi) = \Psi^{-1}(\Psi(u)+\Psi(v))$	$\lim_{u\to 0}\Psi(u) = +\infty$
	Ψ is C^1 and $RV_\alpha(0)$ with $\alpha \in (0,+\infty]$
Gumbel $\Psi(u;\theta) = (-\log(u))^\theta$	$\theta > 1$

Table 1: Examples of copulas satisfying (2.4).

Proposition 2.1 suggests that our identification strategy is natural in the context of Roy's model of self-selection. Following-up on this idea, we consider a generalized Roy model (see Heckman & Vytlacil, 2007, Eisenhauer *et al.* , 2013) with two or more sectors. Let Y_s^* denote the potential outcome corresponding to sector $s \in \{1,...,S\}$, we suppose that

$$Y_s^* = X'\beta_s + (1 + X'\delta_s)\varepsilon_s. \tag{2.5}$$

The utility $\mathcal{U}_s$ associated with sector s is supposed to satisfy

$$\mathcal{U}_s = Y_s^* + G_s(X) + \eta_s. \tag{2.6}$$

The pure Roy model would correspond to $G_s(X) = \eta_s = 0$. Here we also allow for the deterministic and random factors $G_s(X)$ and η_s to affect $\mathcal{U}_s$. Individuals choose the sector that maximizes their utility,

$$D = \arg\max_{s\in\{1,...,S\}} \mathcal{U}_s. \tag{2.7}$$

We further assume a factor structure for the unobservables ε_s and η_s.

Assumption 6. *(i)* $\varepsilon_s = \lambda_{s,1}'\pi + \nu_{s,1}$ *and* $\eta_s = \lambda_{s,2}'\pi + \nu_{s,2}$*, where* π *is a vector of common factors, (ii)* π *has a compact support, (iii)* $(X, \pi, \nu_{1,1}, ..., \nu_{S,2})$ *are mutually independent, (iv)* $\nu_s = \sup_{x\in Supp(X)} \max_{i\neq s}((1+x'\delta_i)\nu_{i,1} + \nu_{i,2} - \nu_{s,2})$ *satisfies the condition of Assumption 3, and (v)* $G_s(X)$ *has compact support.*

The assumption of a linear factor model on the error terms is quite common in the context of generalized Roy models, see e.g. Carneiro *et al.* (2003) or Cunha & Heckman (2007). Proposition 2.2 below shows that, under these conditions, our identification strategy can be used to identify (β_s, δ_s) without any exclusion restriction or large support regressor.[6]

[6]D'Haultfoeuille & Maurel (2013a, 2013b) derived related results, but for a different class of Roy models with only two sectors and in the absence of heteroskedasticity in the outcome equation.

Proposition 2.2. *Suppose that Equations* (2.5)-(2.7) *and Assumptions 2 and 6 hold. Then for all* $s \in \{1, ..., S\}$, $D_s = \mathbb{1}\{D = s\}$ *satisfies Assumption 4, and* (β_s, δ_s) *are identified.*

3 Estimation

3.1 Definition of the estimators

We start by defining our estimators before establishing their asymptotic properties in the next subsection. Suppose we have a sample $(D_i, Y_i, X_i)_{i=1...n}$ of n i.i.d. random variables distributed as (D, Y, X). The starting point for identification is that under Assumptions 1 and 4, we have, as $y \to -\infty$,

$$F_{-Y|X}(y|x) \sim h \, F_{-\varepsilon}((y + x'\beta)/(1 + x'\delta)) \tag{3.1}$$

Now, the key insight for estimating (β, δ) is that if one also imposes Assumption 3, then it is possible to invert both sides and maintain the equivalence. It follows that the quantile regression of $-Y$ on X is asymptotically linear. This result is going to play an important role in our estimation procedure.

Lemma 3.1. *Under Assumptions 1-4, as* $\tau \to 0$,

$$Q_{-Y|X}(\tau|x) \sim \gamma(\tau) + x'\beta(\tau) \tag{3.2}$$

where $\gamma(\tau) = Q_{-\varepsilon}(\tau/h)$ *and* $\beta(\tau) = -\beta + \gamma(\tau)\delta$.

Lemma 3.1 provides the intuition that it might be possible to use quantile regressions in the tails to consistently estimate $(\gamma(\tau), \beta(\tau))$, for small values of the quantile index τ. The main difficulty in formalizing this intuition though, comes from the fact that (3.2) is an equivalence and not an equality, which gives rise to a bias term that needs to be controlled. We define

$$(\widehat{\gamma}(\tau), \widehat{\beta}(\tau)) = \arg\min_{\gamma,\beta} \sum_{i=1}^{n} \rho_\tau(-Y_i - \gamma - X_i'\beta),$$

where $\rho_\tau(u) = (\tau - \mathbb{1}\{u < 0\})u$ is the check function used in quantile regressions. Then one can simply use the following relationships to estimate the parameters of interest β and δ:

$$\begin{aligned} \delta &= \frac{\beta(l\tau) - \beta(\tau)}{\gamma(l\tau) - \gamma(\tau)}, \\ \beta &= -\beta(\tau) + \gamma(\tau)\delta, \end{aligned}$$

where the first equality holds provided that $\gamma(l\tau) - \gamma(\tau) \neq 0$.[7] We basically follow this route in the paper, except that for an efficiency matter we estimate δ using J reduced form estimators $(\beta(l_1\tau), ..., \beta(l_J\tau))$ rather than just two, where $(l_1, \ldots, l_J)$ is a vector of positive spacing parameters such that $l_j \neq 1$ for all $j \in \{1, \ldots, J\}$. Let us consider

$$g_n(\delta) = \begin{pmatrix} \widehat{\beta}(l_1\tau) - \widehat{\beta}(\tau) - (\widehat{\gamma}(l_1\tau) - \widehat{\gamma}(\tau))\delta \\ \vdots \\ \widehat{\beta}(l_J\tau) - \widehat{\beta}(\tau) - (\widehat{\gamma}(l_J\tau) - \widehat{\gamma}(\tau))\delta \end{pmatrix}$$

and let W_n be a $Jd \times Jd$ positive definite symmetric matrix. We estimate δ using a minimum distance procedure:

$$\widehat{\delta} = \arg\min_{\delta} g_n(\delta)' W_n g_n(\delta). \tag{3.3}$$

Finally, we estimate β by averaging across the quantile indices:

$$\widehat{\beta} = \frac{1}{J+1} \sum_{j=0}^{J} -\widehat{\beta}(l_j\tau) + \widehat{\gamma}(l_j\tau)\widehat{\delta},$$

with $l_0 = 1$. We do not simultaneously estimate β and δ since the corresponding estimators of β and δ would have different rates of convergence, thus implying that the standard asymptotic theory of minimum distance estimators would not apply in this context. In particular, this framework would lead to a singular optimal weighting matrix. Intuitively, only the terms with the slowest rate of convergence would be weighted positively, since the other terms would not matter asymptotically. We would then lose consistency of the estimators.

Our estimators depend on a choice of τ, $(l_1, .., l_J)$ and W_n. We derive in the following section the optimal weighting matrix, which can be consistently estimated. Regarding the quantile indices, while the choice of the constants $(l_1, .., l_J)$ does not appear to matter much in practice, an appropriate choice of τ is crucial to balance bias and variance in such a way that guarantees our estimators to be asymptotically normal with zero mean. We propose a data-driven procedure for that purpose in Section 3.3.

We now turn to the estimation of the quantile effects. $\Delta_{j\tau}$ is point identified when $\delta_j = 0$ and partially identified otherwise. Moreover, as shown in Section 3.2, it is possible to estimate β_j more precisely in the case where $\delta_j = 0$. We consider these two cases separately, noting that the restriction $\delta_j = 0$ can be tested using the asymptotic distribution of $\widehat{\delta}_j$ provided in the following section.

[7] Assumptions 3-(i) and 8 below ensure that the latter condition is satisfied for any $l \neq 1$ and τ small enough.

Suppose first that the model is partially homoskedastic, in the sense that $\{\delta_{j_k}\}_{k=1}^{d_\beta}$ are equal to zero for some $d \geq d_\beta \geq 1$ and $\{j_1, ..., j_{d_\beta}\} \subset \{1, ..., d\}$. Then the quantile effects of $\{X_{j_k}\}_{k=1}^{d_\beta}$ correspond to the average effects of the corresponding covariates, which are identified and equal to $\{\beta_{j_k}\}_{k=1}^{d_\beta}$. Let Ψ be a $d_\beta \times d$ matrix that picks out the corresponding subvector of β, i.e. $\beta^1 = \Psi\beta$ and let us consider

$$g_{1n}(\beta^1) = \left(\Psi\widehat{\beta}(\tau) - \beta^1, \Psi\widehat{\beta}(l_1\tau) - \beta^1, ..., \Psi\widehat{\beta}(l_J\tau) - \beta^1\right)'.$$

We then propose to estimate β^1 by

$$\widehat{\beta}^1 = \arg\min_\beta g_{1n}(\beta^1)' W_{1n} g_{1n}(\beta^1),$$

for some positive definite matrix W_{1n}. We also estimate the subvector δ^1 of nonzero components of δ. Letting $\widetilde{\Psi}$ denote the matrix such that $\delta^1 = \widetilde{\Psi}\delta$ and

$$g_{2n}(\delta^1) = \left(\widetilde{\Psi}\left[\widehat{\beta}(l_1\tau) - \widehat{\beta}(\tau)\right] - \left[\widehat{\gamma}(l_1\tau) - \widehat{\gamma}(\tau)\right]\delta^1, ..., \widetilde{\Psi}\left[\widehat{\beta}(l_J\tau) - \widehat{\beta}(\tau)\right] - \left[\widehat{\gamma}(l_J\tau) - \widehat{\gamma}(\tau)\right]\delta^1\right)',$$

we estimate δ^1 by

$$\widehat{\delta}^1 = \arg\min_\delta g_{2n}(\delta^1)' W_{2n} g_{2n}(\delta^1),$$

for some positive definite matrix W_{2n}.

Finally, if we reject partial homoskedasticity so that the quantile effects are only partially identified, one possibility would be to estimate the sharp bounds given in Theorem 2.1. As mentioned previously, however, to the best of our knowledge one cannot conduct inference on these bounds using available methods. Instead, we propose to use the simpler outer bounds given by (2.3). These bounds can be consistently estimated using plug-in estimators, replacing (β, δ) by their estimators $(\widehat{\beta}, \widehat{\delta})$ and $(P(D = 1|X = x), Q_{Y|D=1,X=x})$ by any given consistent nonparametric estimator, e.g., kernel or local polynomial estimators.

3.2 Asymptotic properties

We now turn to the asymptotic properties of $(\widehat{\beta}, \widehat{\delta})$. We rely for that purpose on the asymptotic properties of extremal quantile regressions, established by Chernozhukov (2005). As already discussed, an important difference is that (3.2) is an equivalence rather than an equality. This implies that a bias term comes into play, which needs to be controlled for.

In addition to the previous Assumptions 1-4, our asymptotic analysis relies on the three

conditions discussed below. In the following, we let

$$f(\gamma) = E\left[\sup_{u\leq\gamma} |h - P(D=1|X,-\varepsilon=u)| \times \|\overline{X}\|\right].$$

Assumption 7. *(i.i.d. sampling)* $(D_i, Y_i, X_i)_{i=1...n}$ *are independent, with the same distribution as* (D, Y, X).

Assumption 8. *(Monotone densities) There exists* $A < 0$ *such that almost surely,* $F_{-\varepsilon}$ *and* $F_{-\varepsilon|D=1,X}$ *are differentiable with increasing derivatives on* $(-\infty, A)$.

Assumption 9. *(Rate of convergence of the quantile index)* τ_n *satisfies, as* $n \to \infty$*, (i)* $\tau_n \to 0$*, (ii)* $\tau_n n \to \infty$ *and (iii)* $\sqrt{\tau_n n} f(\gamma(\tau_n)) \to 0$*, where* $\gamma(\tau_n) = Q_{-\varepsilon}(\tau_n/h)$.

Assumption 8 rules out erratic behavior of the densities in the tail. It is very mild and satisfied by all standard distributions. Assumption 9 is an important condition that restricts the rate of convergence of the tail index τ_n. Conditions (i) and (ii) basically ensure that the number of observations that are useful for inference, which is proportional to $\tau_n n$, tends to infinity, but at a slower rate than the sample size. Thus, following the standard terminology in order statistics theory, our estimators are based on quantile regressions where τ_n is an intermediate order sequence, which we will refer to as intermediate order quantile regressions. The reason why we use intermediate order instead of extreme order sequences, where $\tau_n n$ tends to a non-zero constant, is that in the latter case, $\widehat{\delta}$, and thus $\widehat{\beta}$, are not consistent. Intuitively, this is due to the fact that only a finite number of observations are useful in the extreme order case. Intermediate order quantile theory also has the nice feature that it guarantees asymptotic normality rather than convergence towards a non-standard, data-dependent, distribution (see Chernozhukov, 2005 and Chernozhukov & Fernandez-Val, 2011, in the absence of sample selection). Finally, Condition (iii) is specific to our context. This is an undersmoothing condition, which ensures that the bias arising because (3.2) is an equivalence rather than an equality vanishes quickly enough.

Importantly, under Assumption 4, there always exists a τ_n satisfying Assumption 9. Specifically, for any $\alpha \in (0,1)$ define $G(\gamma) = F_{-\varepsilon}(\gamma) f(\gamma)^{2(1-\alpha)}$, where $f(.)$ was introduced at the beginning of the section. By construction, f is increasing. Because $F_{-\varepsilon}$ is strictly increasing on $(-\infty, A)$ by Assumption 8, G is also strictly increasing on $(-\infty, A)$. Then define, for n large enough,

$$\tau_n^* = hF_{-\varepsilon} \circ G^{-1}(1/n). \tag{3.4}$$

Under Assumption 4, $\lim_{\gamma\to-\infty} f(\gamma) = 0$.[8] Thus, $\lim_{\gamma\to-\infty} G(\gamma) = 0$. This implies that $\lim_{n\to\infty} G^{-1}(1/n) = -\infty$, ensuring that τ_n^* satisfies Condition (i). Moreover, it follows from the equality $F_{-\varepsilon}(\gamma) = \frac{G(\gamma)}{f^{2(1-\alpha)}(\gamma)}$ that

$$\tau_n^* = \frac{h/n}{f^{2(1-\alpha)} \circ G^{-1}(1/n)},$$

which implies that Condition (ii) holds as well. Finally, by using this expression again and noting that $\gamma(\tau_n^*) = G^{-1}(1/n)$, we get

$$\sqrt{n\tau_n^*} f(\gamma(\tau_n^*)) = \frac{\sqrt{h} \times f \circ G^{-1}(1/n)}{f^{(1-\alpha)} \circ G^{-1}(1/n)} = \sqrt{h} f^{\alpha} \circ G^{-1}(1/n),$$

so that Condition (iii) is also satisfied. An obvious issue is that such a τ_n^* depends on $F_{-\varepsilon}$ and f, both of which are unknown to the researcher. We shall come back to the issue of the practical choice of τ_n in Section 3.3.

The main result of this section is stated in Theorem 3.1 below, which shows that the estimators of β and δ are consistent and asymptotically normal, and characterizes their asymptotic variances. We first need to introduce several matrices. First, let L be the matrix of typical term $L_{i,j} = \frac{l_{i-1} \wedge l_{j-1}}{\sqrt{l_{i-1} l_{j-1}}}$ for $(i,j) \in \{1, ..., J+1\}^2$. Second, let $\Delta = [-\delta,\ I_d]$, where I_d denotes the identity matrix of size d. Define $\Gamma = [-\iota_J,\ \text{diag}(1/\sqrt{l_1}, ..., 1/\sqrt{l_J})] \otimes I_{d+1}$, where ι_J denotes the column vector of ones of size J and, for any vector v, $\text{diag}(v)$ denotes the diagonal matrix with diagonal v. Finally, let $G = (\log(l_1), ..., \log(l_J))' \otimes I_d$, $\mathcal{Q}_H = E\left[\overline{X_i X_i}'/(1 + X_i'\delta)\right]$ and $\Omega_0 = \mathcal{Q}_H^{-1} \mathcal{Q}_X \mathcal{Q}_H^{-1}$.

Theorem 3.1. *Under Assumptions 1-4 and 7-9, and if $W_n \xrightarrow{p} W$ symmetric positive definite and nonstochastic,*

$$\begin{aligned} \sqrt{\tau_n n}(\widehat{\delta} - \delta) &\xrightarrow{d} \mathcal{N}(0, \Omega_\delta) \\ \frac{\sqrt{\tau_n n}}{\widehat{\gamma}(\tau_n)}(\widehat{\beta} - \beta) &\xrightarrow{d} \mathcal{N}(0, \Omega_\delta) \end{aligned}$$

where $\Omega_\delta = (G'WG)^{-1} G'W(I_J \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta')WG(G'WG)^{-1}$. The optimal weighting matrix is $W_\delta^* = ((I_J \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'))^{-1}$ and the corresponding asymptotic variance is $\Omega_\delta^* = (G'W_\delta^* G)^{-1}$. Finally, there exists τ_n' satisfying Assumption 9 such that
- $\widehat{\beta}$ is consistent if, for some $a > 1$, $f(u) = o(|u|^{-a})$ as $u \to -\infty$.
- The rates of convergence of $\widehat{\delta}$ and $\widehat{\beta}$ are polynomial if for some $a > 0$, $f(u) = o(F_{-\varepsilon}(u)^a)$

[8] To see this, note that for any x, $\sup_{u \le \gamma} |h - P(D = 1|X = x, -\varepsilon = u)|$ tends to zero by Assumption 4. Because this term is bounded by 2, f tends to zero by the dominated convergence theorem.

as $u \to -\infty$.[9]

A consistent estimator of the asymptotic variance Ω_δ can be obtained by replacing W by W_n, Δ by $\widehat{\Delta} = [-\widehat{\delta},\ I_d]$ and Ω_0 by $\widehat{\Omega}_0 = \widehat{\mathcal{Q}}_H^{-1}\widehat{\mathcal{Q}}_X\widehat{\mathcal{Q}}_H^{-1}$, with

$$\widehat{\mathcal{Q}}_X = \frac{1}{n}\sum_{i=1}^n \overline{X_i X_i}', \quad \widehat{\mathcal{Q}}_H = \frac{1}{n}\sum_{i=1}^n \overline{X_i X_i}'/(1 + X_i'\widehat{\delta}).$$

Similarly, one can consistently estimate W_δ^*, and thus obtain a two-step estimator that is optimal in the class of estimators considered here.

Theorem 3.1 shows that $\widehat{\delta}$ converges more quickly than $\widehat{\beta}$ towards the true value, since $|\gamma(\tau_n)| \to \infty$ as $\tau_n \to 0$. Actually, even though one can always construct asymptotically valid confidence intervals on β, $\widehat{\beta}$ may not be consistent. Consistency is secured, however, if f decreases to zero quickly enough. To understand what this condition means, it is useful to discuss it in the context of the sample selection model defined by Assumption 5. In such a case, letting $\overline{x} = [1, x']'$, we have

$$\begin{aligned}
f(\gamma(\tau)) &\leq \sup_{\overline{x}\in \mathrm{Supp}(X)} ||\overline{x}|| \sup_{x\in\mathrm{Supp}(X), u\leq\gamma(\tau)} |P(D=1|X=x,-\varepsilon=u)-1| \\
&= \sup_{\overline{x}\in \mathrm{Supp}(\overline{X})} ||\overline{x}|| \sup_{x\in\mathrm{Supp}(X), u\leq\gamma(\tau)} |P(\eta\leq\phi(x)|-\varepsilon=u)-1| \\
&= \sup_{\overline{x}\in \mathrm{Supp}(\overline{X})} ||\overline{x}|| \sup_{x\in\mathrm{Supp}(X), u\leq\gamma(\tau)} |P(F_\eta(\eta)\leq F_\eta(\phi(x))|F_{-\varepsilon}(-\varepsilon)=F_{-\varepsilon}(u))-1| \\
&\leq \sup_{\overline{x}\in \mathrm{Supp}(\overline{X})} ||\overline{x}||\ f_C(\tau),
\end{aligned}$$

where the second inequality follows from $f_C(\tau) = \sup_{u\leq\tau, v\in[\underline{v},1]} |\partial_1 C(u,v) - 1|$. Hence, consistency of $\widehat{\beta}$ is achieved if $f_C(\tau) = o(|\gamma(\tau)|^{-a})$ for some $a > 1$. Similarly, if, for some $b > 0$,

$$f_C(\tau) = o(\tau^b), \tag{3.5}$$

then a polynomial rate of convergence, faster than $n^{(b-\zeta)/(2b+1)}$ for any $b > \zeta > 0$, is possible. Table 2 below provides examples of copulas of $(\eta, -\varepsilon)$ satisfying the latter condition (see Appendix C.7 for its verification in each case), and therefore copulas for which $\widehat{\beta}$ is consistent.[10] It is worth noting that for the last two copulas considered in the table, we actually establish that $f_C(\tau)$ tends to zero exponentially fast in τ. In such situations, (3.5) holds for all b, and

[9] Assumption 3 implies that for all $a > 0$, $F_{-\varepsilon}(u) = o(|u|^{-a})$. Thus, the condition $f(u) = o(F_{-\varepsilon}(u)^a)$ is stronger than the one ensuring consistency of $\widehat{\beta}$, as expected.

[10] See, e.g., Nelsen (2006) for a detailed review of copulas and their properties.

it is possible to achieve a rate of convergence for $\widehat{\delta}$ and $\widehat{\beta}$ that is faster than $n^{1/2-\zeta}$ for any $\zeta > 0$. In other words, an adequate choice of τ_n can make the rate of convergence arbitrarily close to the standard parametric root-n rate.

Copula family	Restriction ensuring (3.5)
Gaussian $C(u, v; \rho)$	$\rho > 0$
Clayton $C(u, v; \theta) = \max\left([u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0\right)$	$\theta > 0$
Rotated Gumbel-Barnett $C(u, v; \theta) = u - u(1-v)\exp(-\theta \log(u)\log(1-v))$	$\theta \in (0, 1]$
$C(u, v; \theta) = (1 + [(u^{-1} - 1)^{\theta} + (v^{-1} - 1)^{\theta}]^{1/\theta})^{-1}$	$\theta > 1$
$C(u, v; \theta) = (1 + [(u^{-1/\theta} - 1)^{\theta} + (v^{-1/\theta} - 1)^{\theta}]^{1/\theta})^{-\theta}$	$\theta \geq 1$
$C(u, v; \theta) = \theta / \log(\exp(\theta/u) + \exp(\theta/v) - \exp(\theta))$	$\theta > 0$
$C(u, v; \theta) = [\log(\exp(u^{-\theta}) + \exp(v^{-\theta}) - e)]^{-1/\theta}$	$\theta > 0$

Table 2: Examples of copulas leading to a polynomial rate of convergence.

In order to conduct inference on the quantile effects, we need to distinguish between the partially homoskedastic case and the heteroskedastic case. This involves (pre)testing the restriction $\delta_j = 0$. Valid inference requires that the critical value of the corresponding t-test depend on the sample size n, so that the level of the test tends to zero while the power tends to one. A possibility is to choose the critical values c_n so that $c_n \to \infty$, but slowly enough so that $c_n/\sqrt{n} \to 0$. In practice, we use in our application $c_n = \sqrt{\log(n)}$, which is advocated in different contexts by Andrews (1999) and Andrews & Soares (2010).

We first consider the partially homoskedastic case. As before, we let $G_\delta = (\log(l_1), ..., \log(l_J))' \otimes I_{d-d_\beta}$, $G_\beta = -\iota_{J+1} \otimes I_{d_\beta}$, $\Gamma_2 = (0, I_d)$ and $\Gamma_3 = \text{diag}(1, 1/\sqrt{l_1}, \cdots, 1/\sqrt{l_J})$. Finally, we let $\widehat{\lambda}_n = \widehat{\gamma}(\tau_n)\log(m)/(\widehat{\gamma}(m\tau_n) - \widehat{\gamma}(\tau_n))$, where $m \neq 1$ denotes an arbitrary positive constant.

Theorem 3.2. *Under Assumptions 1-4 and 7-9, $\{\delta_{i_k}\}_{k=1}^{d_\beta}$ are zeros and if $W_{1n} \xrightarrow{p} W_1$ and $W_{2n} \xrightarrow{p} W_2$, where W_1 and W_2 are symmetric positive definite and nonstochastic, then*

$$\begin{aligned} \sqrt{\tau_n n}(\widehat{\delta}^1 - \delta^1) &\xrightarrow{d} \mathcal{N}\left(0, \Omega_{\delta^1}\right) \\ \widehat{\lambda}_n \frac{\sqrt{\tau_n n}}{\widehat{\gamma}(\tau_n)}(\widehat{\beta}^1 - \beta^1) &\xrightarrow{d} \mathcal{N}\left(0, \Omega_{\beta^1}\right), \end{aligned}$$

where

$$\begin{aligned}\Omega_{\delta^1} &= \left[G_\delta' W_2 G_\delta\right]^{-1} G_\delta' W_2 (I_J \otimes \widetilde{\Psi}\Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'\widetilde{\Psi}')W_2 G_\delta \left[G_\delta' W_2 G_\delta\right]^{-1},\\ \Omega_{\beta^1} &= (G_\beta' W_1 G_\beta)^{-1} G_\beta' W_1 (\Gamma_3 \otimes \Psi\Gamma_2)(L \otimes \Omega_0)(\Gamma_3' \otimes \Gamma_2'\Psi')W_1 G_\beta (G_\beta' W_1 G_\beta)^{-1}.\end{aligned}$$

The optimal weighting matrices for $\widehat{\delta_1}$ *and* $\widehat{\beta_1}$ *are then* $W^*_{\delta^1} = \left[(I_J \otimes \widetilde{\Psi}\Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'\widetilde{\Psi}')\right]^{-1}$ *and* $W^*_{\beta^1} = [(\Gamma_3 \otimes \Psi\Gamma_2)(L \otimes \Omega_0)(\Gamma_3' \otimes \Gamma_2'\Psi')]^{-1}$*, and the corresponding asymptotic variances are given by* $(G_\delta' W^*_{\delta^1} G_\delta)^{-1}$ *and* $(G_\beta' W^*_{\beta^1} G_\beta)^{-1}$*. Finally,* $|\widehat{\lambda}_n| \xrightarrow{p} \infty$*, so that the rate of convergence of* $\widehat{\beta^1}$ *is faster than the one of the unconstrained estimator* $\widehat{\beta}$*.*

Finally, in the case where we reject $\delta_j = 0$, we propose to construct a confidence interval on the quantile effect $\Delta_{j\tau}$ based on the outer bounds given by (2.3). Importantly, because the quantities $\underline{Q}^o_\varepsilon(\tau)$ and $\overline{Q}^o_\varepsilon(\tau)$ are defined as supremum and infimum of functions that have to be estimated, we can apply the methodology developed by Chernozhukov *et al.* (2013) to conduct inference on intersection methods. More details on the construction of these confidence intervals in our context are provided in Appendix A.1.

3.3 Choice of the quantile index

The estimators of β and δ are asymptotically normal with zero mean provided that they are based on a sequence of quantile indices τ_n satisfying the bias-variance trade-off of Assumption 9. Though there always exists a sequence τ_n satisfying Assumption 9 under Assumption 4, admissible rates of convergence towards 0 for τ_n are unknown, since they depend on $f(\gamma(\tau_n))$, which is itself unknown. A related issue arises in the estimation at infinity of the intercept of sample selection models (see Andrews & Schafgans, 1998 and Schafgans & Zinde-Walsh, 2002) or in the estimation of extreme value index (see Drees & Kaufmann, 1998 and Danielsson *et al.* , 2001). We propose in the following a heuristic data-driven method, which consists of selecting τ_n as the minimizer of a criterion function that represents the trade-off between bias and variance. The innovative idea here is to combine a subsampling method with a minimum distance estimator to produce a proxy of the bias.

Specifically, let us consider the J test statistic

$$T_J(\tau) = \frac{\log(m)^2 \tau n}{\left(\widehat{\gamma}(m\tau) - \widehat{\gamma}(\tau)\right)^2} g_n(\widehat{\delta}(\tau))' \widehat{W}^*_\delta g_n(\widehat{\delta}(\tau)),$$

for some arbitrary fixed $m > 1$. Here $\hat{\delta}(\tau)$ is the estimator obtained using the quantile index τ and $\widehat{W}^*_\delta$ is an estimator of W^*_δ using $\hat{\delta}(\tau)$. We prove in Appendix A.2 that if τ_n satisfies

Assumption 9, $T_J(\tau_n)$ converges to a chi-square distribution with $(J-1)d$ degrees of freedom as n grows to infinity. We also show that otherwise, the asymptotic distribution of the J test statistic includes an additional term. Heuristically, this suggests in particular that if the median of the J test statistic is close enough to the median of a chi-square distribution with $(J-1)d$ degrees of freedom, denoted by $M_{(J-1)d}$, then the bias term should be small. Our data-driven procedure builds on this idea.

In practice, we propose to estimate the difference between the two medians using subsampling. For each subsample and each quantile index τ within a grid defined below, we compute $T_J(\tau)$. Then, letting $M_s(\tau)$ denote the median of these test statistics over the different subsamples and for a given τ, we compute

$$\widehat{\text{diff}}_n(\tau) = \frac{\left|M_s(\tau) - M_{(J-1)d}\right|}{\sqrt{b_n \tau}},$$

where b_n denotes the subsample size.

Similarly, the asymptotic variance is estimated by the variance of the subsampling point estimates of δ multiplied by the normalizing factor b_n/n. We call this estimator $\widehat{\text{Var}}_n(\tau)$. At the end, we select the quantile index as follows:

$$\widehat{\tau}_n = \arg\min_{\tau} \widehat{\text{Var}}_n(\tau) + \widehat{\text{diff}}_n(\tau).$$

We thus base our procedure on the trade-off between the variance and our proxy of the bias. It follows that we achieve undersmoothing in comparison with a more standard trade-off between variance and squared bias. Note that, similarly to the case of nonparametric regressions, this is needed to control the asymptotic bias that would otherwise affect the limiting distribution of our estimator.

We implement this method by searching over a grid of τ on an interval. In practice, we set the upper bound of this interval to 0.3 and the lower bound to $80/b_n$. This lower bound is motivated by the fact that if the effective subsampling size τb_n becomes too small, then the intermediate order asymptotic theory is likely to be a poor approximation (see Chernozhukov & Fernandez-Val, 2011 for a related discussion). Finally, we select the quantile indices for estimating β and δ in the partially homoskedastic case in the same manner.

4 Simulations

In this section, we investigate the finite-sample performances of our estimation procedure by simulating the following model for four different sample sizes ($n = 250$, $n = 500$, $n = 1,000$ and $n = 2,000$):

$$\begin{aligned} Y^* &= \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + (1 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3)\varepsilon \\ D &= \mathbb{1}\left\{0.6 + Y^* + 0.3X_1 + 0.2X_2 + X_3^2 + \eta \geq 0\right\}. \end{aligned}$$

X_1 and X_2 are two mutually exclusive binary variables, such that $X_1 = \mathbb{1}\{U \leq 0.3\}$ and $X_2 = \mathbb{1}\{U \geq 0.8\}$, with U uniformly distributed over $[0, 1]$. X_3 is drawn from a truncated normal distribution with support $[-1.8, 1.8]$, mean 0 and standard deviation 1. (ε, η) are jointly normally distributed, with mean zero and covariance matrix $\begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$. Finally, the true values of the parameters are given by: $\beta_1 = 0.2$, $\beta_2 = 0.4$, $\beta_3 = 0.5$, $\delta_1 = 0$, $\delta_2 = 0.1$ and $\delta_3 = -0.3$.

We report in Table 3 below, for each sample size, the bias and standard deviation for nine different estimators. Namely, we first estimate $(\delta_1, \delta_2, \delta_3)$ and $(\beta_1, \beta_2, \beta_3)$ without imposing $\delta_1 = 0$. Then we impose $\delta_1 = 0$ and reestimate β_1 along with (δ_2, δ_3) under this partial homoskedasticity constraint. As shown in Section 3.2 (Theorem 3.2), β_1 is estimated at a faster convergence rate in the latter case. In both cases, we use the two-step, asymptotically optimal estimators of β and δ. We also document the severity of the selection bias in this context by reporting the bias and standard deviation of a naive OLS estimator of β_1 only using the observations such that $D = 1$. Throughout this section we pay special attention to the performances of our constrained estimator of β_1 since, in our application, the black-white wage gap will be estimated similarly.

The vector of spacing parameters l_j used in minimum distance estimation is set equal to (0.65, 0.85, 1.15, 1.45). Intuitively, these parameters have to differ sufficiently to provide enough variation. At the same time, they should not be too large, otherwise the corresponding quantile indices $\tau_n l_j$ might escape from the extremal quantiles region. However, in practice, our estimates do not appear to be meaningfully sensitive to the choice of l. The choice of the quantile index τ_n is more critical. We choose this parameter using the data-driven method discussed in Section 3.3, with subsample sizes $(150, 300, 500, 600)$ corresponding to the four total sample sizes (250, 500, 1,000 and 2,000) and 500 subsamples in each case. We report in Table 3 below the average quantile indices computed across all simulations.

	Heteroskedastic						Homoskedastic ($\delta_1 = 0$)			OLS
	δ_1	δ_2	δ_3	β_1	β_2	β_3	δ_2	δ_3	β_1	β_1
True Value	0	0.1	-0.3	0.2	0.4	0.5	0.1	-0.3	0.2	0.2
n=250										
Bias	0.070	0.104	0.086	-0.018	-0.053	-0.054	0.066	0.065	0.021	-0.075
Std dev	0.305	0.395	0.148	0.252	0.318	0.099	0.430	0.154	0.187	0.152
Average τ_n		0.256			0.256			0.236	0.207	
n=500										
Bias	0.073	0.074	0.064	-0.041	-0.051	-0.040	0.012	0.053	0.012	-0.076
Std dev	0.260	0.358	0.128	0.208	0.283	0.098	0.334	0.124	0.137	0.102
Average τ_n		0.220			0.220			0.209	0.201	
n=1,000										
Bias	0.023	0.025	0.031	-0.018	-0.013	-0.019	0.004	0.032	-0.010	-0.078
Std dev	0.192	0.230	0.082	0.176	0.211	0.069	0.241	0.083	0.089	0.072
Average τ_n		0.203			0.203			0.201	0.208	
n=2,000										
Bias	0.020	0.045	0.020	-0.009	-0.035	-0.015	0.008	0.011	0.000	-0.077
Std dev	0.134	0.192	0.064	0.126	0.171	0.055	0.175	0.051	0.062	0.054
Average τ_n		0.191			0.191			0.185	0.203	

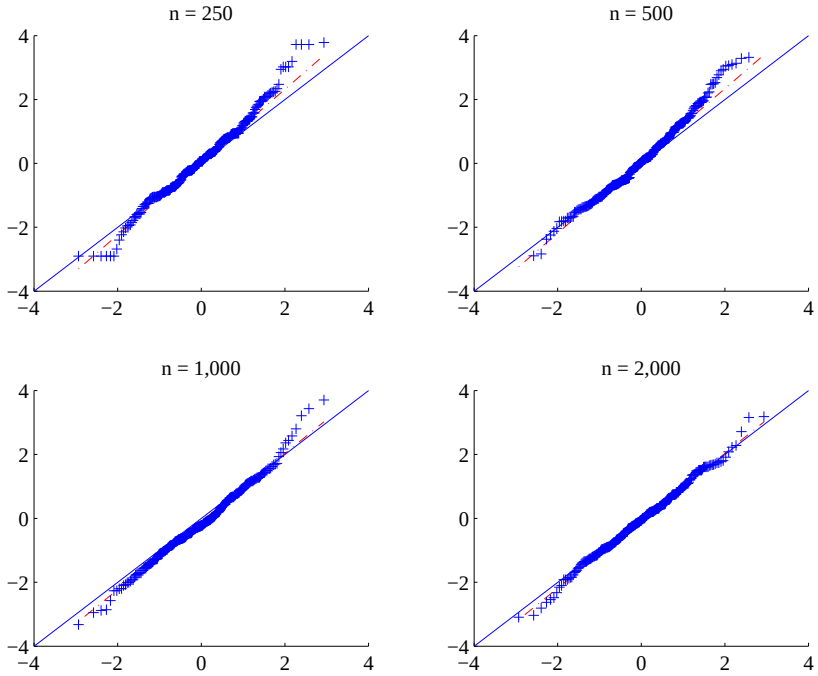
Note: Results were obtained using 300 simulations for each sample size.

Table 3: Monte Carlo simulations

Importantly, for each sample size, the bias-standard deviation ratio for each estimator is much smaller than 1, consistent with our data-driven choice of τ_n leading to undersmoothing. Besides, the standard deviations of our estimators as well as the average τ_n generally decrease with the sample size, as expected given the consistency of our estimators and the bias-variance tradeoff underlying the choice of τ_n.[11] In practice, our estimators exhibit a fairly small bias for sample sizes larger than $n = 500$. Note also that the constrained version of the estimator of β_1, which makes use of the partial homoskedasticity constraint $\delta_1 = 0$, is much more precise and yields a smaller bias (except for $n = 250$) than the unconstrained estimator. The OLS estimator of β_1, on the other hand, displays a large bias, which remains stable across all sample sizes.

We report in Figure 1 below the QQ-plots of β_1, after imposing homoskedasticity ($\delta_1 = 0$), for all four sample sizes. The plots are generally close to the diagonal line, which shows that the estimator of β_1 is approximately normally distributed, even for small sample sizes. Importantly, this provides evidence that it is in practice reasonable, even in small samples, to conduct inference based on the asymptotic distributions of our estimators.

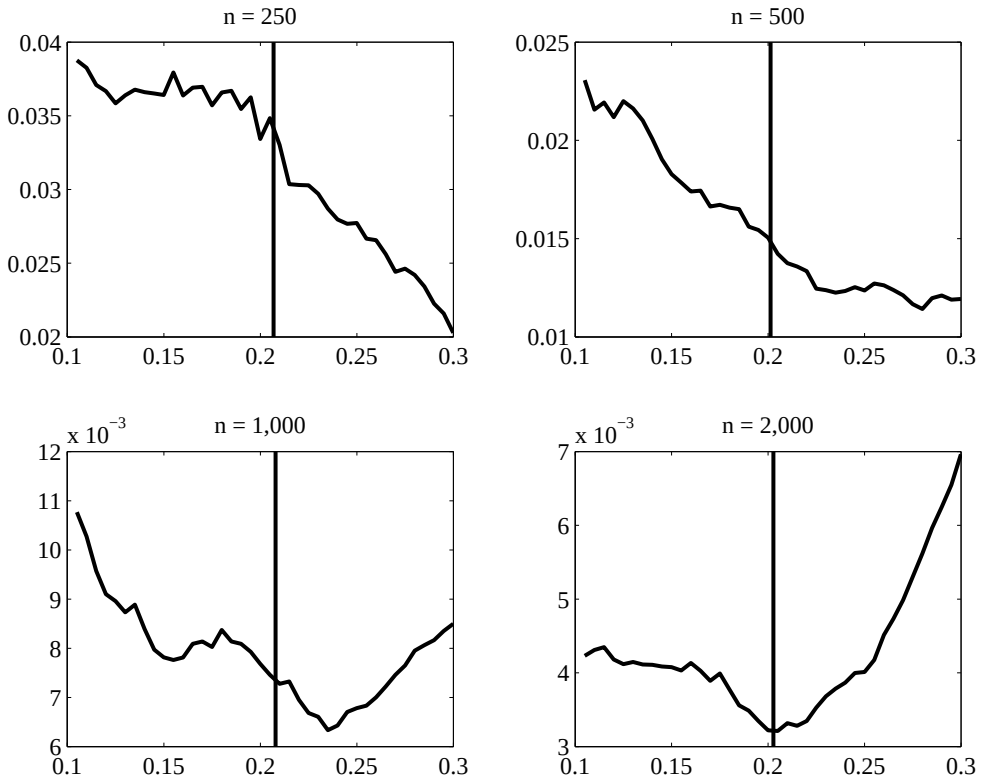
[11] An exception to the overall decrease in τ_n with the sample size is the slight increase of τ_n between $n = 500$ and $n = 1,000$ for the constrained estimator of β_1. τ_n is essentially stable across all sample sizes for this estimator.



Note: The crosses represent the quantiles of the standard normal distribution, against the quantiles of the empirical distribution of the studentized estimates of β_1.

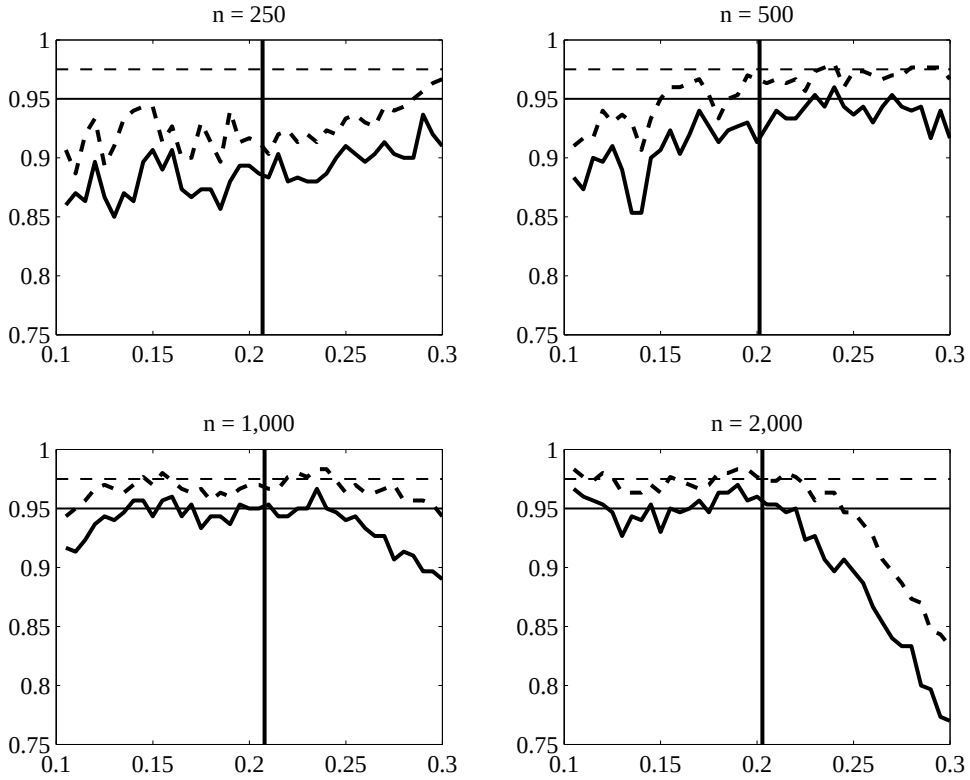
Figure 1: QQ-plots (constrained estimator of β_1)

Figure 2 below shows the evolution of the Mean Squared Error (MSE) of the constrained estimator of β_1, with respect to the quantile index τ_n. The vertical line corresponds to the average τ_n (across simulations) chosen based on our data-driven method. The plots corresponding to $n = 1,000$ and $n = 2,000$ exhibit a U-shaped relationship between the MSE and the quantile index. This pattern reflects a bias-variance tradeoff with respect to the choice of τ_n. When the quantile index is small, the bias is small but the variance is large, and vice versa. On the other hand, the relationship between MSE and τ_n is mostly decreasing for $n = 250$ and $n = 500$. This is consistent with the variance term dominating the bias term for $\tau_n < 0.3$ and such small sample sizes. For all sample sizes, the average quantile index selected with our method is generally smaller than the one yielding the smallest MSE, consistent with our data-driven method tending to undersmooth. However, for sample sizes larger than 500, it is worth noting that the MSE evaluated at the average selected quantile index is close to the minimum (even equal for $n = 2,000$).



Note: The solid vertical line is the average quantile index produced by our data-driven method. The solid curve plots the MSE of our estimator as a function of the quantile index τ_n.

Figure 2: Relationship between MSE (Y-axis) and τ_n (X-axis), constrained estimator of β_1



Note: The solid vertical line is the average quantile index produced by our data-driven method. The horizontal dashed and solid lines represent the 97.5% and 95% nominal coverage rates, respectively. The dashed and solid curves represent the coverages of the 95% and 97.5% confidence intervals, as a function of the quantile index τ_n.

Figure 3: Relationship between coverage (Y-axis) and τ_n (X-axis), constrained estimator of β_1

Finally, we examine in Figure 3 the relationship between the coverage of the 95% and 97.5% confidence intervals constructed with our constrained estimator of β_1 and the quantile index τ_n. The confidence intervals are constructed using the asymptotic variance of our estimator together with normal critical values.[12] The coverage gets reasonably close to the nominal rates for $n = 500$, and remarkably close for larger sample sizes, for values of τ_n in the neighborhood of the average quantile index obtained with our data-driven method. The sharp decline in coverage for large values of the quantile index for $n = 1,000$ and $n = 2,000$ reflects the existence of a nonvanishing bias. For smaller sample sizes, in particular for $n = 500$, the coverage decreases for small values of the quantile index. This may be due to the fact that the quantile index falls in the extreme, rather than intermediate order region, in which case the finite sample distribution of our estimator cannot be approximated well by a standard

[12] We also computed confidence intervals based on percentile bootstrap or subsampling. Overall, the method based on normal critical values performed best in terms of coverage.

normal distribution.As sample size increases, however, the extreme order region moves even closer to the origin, and for all the values of τ_n we consider the sampling distribution remains close enough to a normal distribution. Importantly, for all sample sizes except $n = 250$, the quantile index obtained from our data-driven procedure appears to be relatively close to optimal, in the sense of minimizing the discrepancy between the actual and nominal coverages. For the smallest sample size ($n = 250$) though, our procedure yields a value of the quantile index that appears to be suboptimal, both in terms of coverage and MSE (see Figure 2 above). This is presumably due to the fact that, in this context, the variance term strongly dominates the bias term. In any case, the results reported in Table 3 are quite encouraging since they show that, even with such a small sample, the bias and variance of our constrained estimator of β_1 remain reasonable.

5 Application to the black-white wage gap

We apply our method to the estimation of the black-white wage gap among young males for two groups of cohorts, using data from the National Longitudinal Survey of Youth 1979 (NLSY79) and National Longitudinal Survey of Youth 1997 (NLSY97). Individuals surveyed in the NLSY79 were 14 to 22 years old in 1979, while individuals from the NLSY97 were 12 to 16 years old in 1997. In the following, we are interested in estimating the black-white wage gap for these two groups of individuals as of 1990-1991 and 2007-2008, respectively. As noted in early articles by Butler & Heckman (1977) and Brown (1984), and documented more recently by Juhn (2003), among males, blacks are significantly more likely to dropout from the labor market. To the extent that those dropouts tend to have lower potential wages, it follows that failure to control for endogenous labor market participation is likely to result in underestimating the black-white wage differential. It is worth noting that finding a valid instrument for selection is particularly difficult in the context of male labor force participation. As a result, most of the attempts to deal with selection have consisted of imputing wages for non-workers (see, among others, Brown, 1984, Smith & Welch, 1989, Neal & Johnson, 1996, Juhn, 2003, Neal, 2004, Neal, 2006 and Neal & Rick, 2014).

Importantly, since across-cohort changes in selection into the workforce is also different for blacks and for whites, adequately dealing with selection is needed to obtain credible estimates of the across-cohort evolution of the black-white wage gap. Altonji & Blank (1999) stress the importance of correcting for changes in race differential selection into work, and review some of the empirical literature addressing this issue.[13]

[13] As the authors put it, "Comparisons of average or median wages of persons with jobs do not provide an

5.1 Evidence from the NLSY79

We first use our method to estimate the black-white wage gap among young males from the NLSY79, revisiting the influential work of Neal & Johnson (1996) on this question. We use the same sample as Neal & Johnson (1996) in our analysis, and consider as they did that an individual is a nonparticipant if he did not work in 1990 nor in 1991. The total sample size is $n = 1,674$, with an overall labor force participation rate over the period of interest (1990-1991) equal to 95%. We refer the reader to Neal & Johnson (1996) for a detailed discussion on the data.

We start by replicating the results of Neal & Johnson (1996) in Table 4 below by running four regressions on the log of hourly wages on a set of observable characteristics, namely black, Hispanic dummies and age (specifications (1) and (3)), together with AFQT and AFQT squared (specifications (2) and (4)). The first two columns contain the results of simple OLS regressions, replicating Columns (1) and (3) in Table 1 of Neal & Johnson (1996) (p.875), while in the last two columns we replicate their Table 4 (p.883) by imputing a zero log-wage for nonparticipants and running a median log-wage regression. As discussed in Neal & Johnson (1996) and more extensively in Johnson *et al.* (2000), this imputation method yields consistent estimates under the assumption that, conditional on the set of observable characteristics included in the regression, the potential wage for any individual who did not work neither in 1990 nor in 1991 lies below the median. It is important to note that the identifying condition of independence at infinity used in our paper (Assumption 4) relaxes this assumption by replacing the median with some extremal quantile of the conditional wage distribution.[14] As is put forward by Neal & Johnson (1996), Columns (1) and (2) show that the estimated black-white wage gap drops sharply, from 24.4% to 7.1%, after adding controls for ability, namely AFQT and AFQT squared. It is also worth noting that the estimated black-white wage differential changes substantially, increasing (in absolute value) by as much as 6.4 points, after addressing the selection issue with the imputation method proposed in Neal & Johnson (1996) (see Columns (2) and (4)).

accurate picture of changes in the offer distributions faced by black and by white workers" (pp. 3240). See also Juhn (2003), who provides evidence that the evolution over the period 1969-1998 of the black-white wage gap is severely biased if one does not take into account the decline in work participation rates of black men relative to white men. In recent work, Neal & Rick (2014) show that the growth in prison populations in the last decades is an important factor behind the evolution of differential workforce participation of blacks and whites.

[14] Our identifying condition is also weaker in the sense that h does not need to be equal to 1.

	(1)	(2)	(3)	(4)
Black	-0.244	-0.071	-0.356	-0.135
	(0.026)	(0.027)	(0.028)	(0.034)
Hispanic	-0.114	0.005	-0.181	-0.013
	(0.030)	(0.030)	(0.033)	(0.038)
Age	0.048	0.040	0.068	0.055
	(0.014)	(0.013)	(0.016)	(0.017)
AFQT	___	0.173	___	0.206
		(0.012)		(0.015)
$AFQT^2$	___	-0.013	___	0.010
		(0.011)		(0.014)

Note: Standard errors are reported in parentheses.

Table 4: OLS and median log-wage regression results (NLSY79)

We now investigate how the above results are changed when we use our estimation method and implement the two-step asymptotically optimal estimators of δ and β. Table 5 presents the estimation results for the heteroskedasticity parameters δ and the parameters β. Since we fail to reject homoskedasticity for all the covariates with the exception of age, we report both the corresponding unconstrained ("Heteroskedastic") and constrained ("Homoskedastic") estimates of β. In the discussion below we focus on our preferred constrained estimates, which have a structural interpretation in terms of average effects.

	Heteroskedastic		Homoskedastic
	δ	β	β
Black	0.019	-0.215	-0.101
	(0.029)	(0.326)	(0.010)
Hispanic	0.005	0.014	0.020
	(0.030)	(0.337)	(0.012)
Age	-0.029	0.215	___
	(0.003)	(0.032)	
AFQT	-0.005	0.238	0.215
	(0.012)	(0.134)	(0.005)
$AFQT^2$	-0.007	0.030	0.000
	(0.011)	(0.121)	(0.004)

Note: Standard errors are reported in parentheses. We perform the homoskedasticity tests using the critical values $c_n = \sqrt{\log(n)}$, where n is the sample size ($n = 1,674$ here). The vector of spacing parameters l_j used in minimum distance estimation is equal to $(0.65, 0.85, 1.15, 1.45)$. The quantile index τ_n is chosen based on the data-driven procedure discussed in Section 3.3, using 500 subsamples of size 550.

Table 5: Extremal quantile regression results (NLSY79)

The estimation results from our extremal quantile method show that the size of the black-white wage gap (10.1%) is smaller than the estimated gap obtained under the imputation method proposed by Neal & Johnson (1996) (13.5%), but larger than the gap estimated using simple OLS (7.1%). The fact that our preferred estimate of the black-white wage gap is smaller than the one obtained with the imputation method is consistent with our estimator being based on a weaker identifying assumption. While Neal & Johnson (1996) assume that, conditional on observed characteristics, those individuals who do not participate to the labor market have a potential wage below the median, a sufficient condition to apply our method is to rule out the possibility that non-participants have arbitrarily large potential wages. Intuitively it follows that our approach results in a milder form of selection correction, which is consistent with our findings.

Finally, it is worth stressing that our results are in line with the key takeaway of Neal & Johnson (1996), namely that premarket factors, as measured here by AFQT, account for most of the black-white wage differential. In fact, our results point to an even more important role played by AFQT, since the estimated wage gap drops from close to the median regression estimate (around 35%) to 10.1% after adding AFQT.[15]

5.2 Across-cohort evolution

We now examine the evolution across the NLSY79 and NLSY97 cohorts of the black-white wage gap. To do so, we apply our method to estimate the wage gap using hourly wages measured in 1990-1991 for the NLSY79 sample and in 2007-2008 for the NLSY97 sample. We follow Altonji *et al.* (2012) by using a modified version of the AFQT variable, which corrects for the across-cohort changes in the ASVAB test format as well as in the age ranges at which the test was taken. This age correction procedure is based on an equipercentile mapping. To the extent that the rank within the AFQT distribution may vary with the age of the respondent at the time of the test, we further restrict the samples to the respondents who took the test when they were 16 or 17. Besides this age restriction, we constructed the NLSY97 sample so as to match as closely as possible the sample selection rules used by Neal & Johnson (1996) for the NLSY79. Consistent with prior evidence, we find that the labor force participation rate of black men has fallen over time relative to white men (see Appendix B for more details on the data). The baseline estimation results are reported in Table 6 below. The resulting sample sizes are equal to $1,077$ and $1,123$ for the NLSY79 and NLSY97 cohorts, respectively.

[15] Estimation results from our method without controlling for AFQT are not reported here to save space. They are available from the authors upon request.

	NLSY79		NLSY97	
	Extremal Quantile	Median	Extremal Quantile	Median
Black	-0.122	-0.145	-0.140	-0.167
	(0.001)	(0.039)	(0.050)	(0.058)
Hispanic	0.029	-0.017	-0.054	-0.089
	(0.002)	(0.056)	(0.050)	(0.050)
AFQT	0.185	0.180	0.153	0.111
	(0.001)	(0.019)	(0.022)	(0.026)
$AFQT^2$	0.007	0.008	0.002	-0.023
	(0.001)	(0.017)	(0.022)	(0.020)

Notes: Estimations also include linear control for age. Standard errors are reported in parentheses. In the column "Extremal Quantile", we report the results corresponding to our preferred constrained specification, since we fail to reject homoskedasticity for all of the covariates with the exception of age. We perform the homoskedasticity tests using the critical values $c_n = \sqrt{\log(n)}$, where n is the sample size. The vector of spacing parameters l_j used in minimum distance estimation is equal to $(0.65, 0.85, 1.15, 1.45)$. The quantile index τ_n is chosen based on the data-driven procedure discussed in Section 3.3, using 500 subsamples of size 500.

Table 6: Extremal quantile and median regression results (NLSY79-NLSY97)

The estimation results obtained with our method ("Extremal Quantile" columns) provide evidence of a *wider* black-white wage gap for the 1997 cohort relative to the 1979 cohort, with an increase in the estimated gap from 12.2% to 14%. It is also interesting to note that, while the estimated levels do differ across both methods, the results from the median regression of Neal & Johnson (1996) ("Median" columns) imply an across-cohort increase of a similar magnitude (from 14.5% to 16.7%).

It is important to step back and try and understand what these results really mean. Specifically, do they suggest that labor market discrimination against blacks has actually gotten worse over the last two decades? Or does the estimated increase in the black-white wage gap reflect the fact that the AFQT score only captures a fraction of all the premarket factors that matter on the labor market, which may have changed over time? In particular, the results reported in Table 6 provide clear evidence of a decline across cohorts in the wage returns to AFQT, consistent with the latter story. Recent work by Castex & Dechter (2014) also provides evidence from the NLSY79 and NLSY97 that the wage returns to AFQT have decreased over time (see also Beaudry *et al.* , 2013, who argue that there has been a decline in the demand for cognitive skills in the U.S. since 2000). While providing a definite answer to those questions is particularly challenging, we attempt to shed light on this issue by controlling for additional premarket factors, namely parental education and household structure (as measured by the presence of both biological parents at age 14). Bringing those characteristics into the analysis

is important since differences in family environment have been found to account for most of the black-white gap in noncognitive skills (see, e.g., Carneiro *et al.* , 2005).

Table 7 below reports the estimated black-white wage gap for the 1979 and 1997 cohorts, using our extremal quantile method and the median regression of Neal & Johnson, for three different specifications. The first specification ("No premarket factors") only controls for age and the Hispanic dummy, the second specification ("AFQT only") also controls for AFQT and AFQT squared, while the third specification ("Preferred") further controls for parental education and household structure.

	NLSY79		NLSY97	
	Extremal Quantile	Median	Extremal Quantile	Median
Black (No premarket factor)	-0.342	-0.349	-0.296	-0.311
	(0.014)	(0.032)	(0.003)	(0.051)
Black (AFQT only)	-0.122	-0.145	-0.140	-0.167
	(0.001)	(0.039)	(0.050)	(0.058)
Black (Preferred)	-0.099	-0.123	-0.087	-0.135
	(0.017)	(0.042)	(0.043)	(0.064)

Notes: Standard errors are reported in parentheses. The "preferred" specification includes AFQT, parental education and household structure. For that case, the sample is then restricted to the individuals with non-missing parental education and household structure, resulting in sample sizes equal to $1,016$ for the NLSY79 and $1,071$ for the NLSY97. In the column "Extremal Quantile", we report the results corresponding to our preferred constrained specification, since we fail to reject homoskedasticity for the black dummy. We perform the homoskedasticity tests using the critical values $c_n = \sqrt{\log(n)}$, where n is the sample size. The vector of spacing parameters l_j used in minimum distance estimation is equal to $(0.65, 0.85, 1.15, 1.45)$. The quantile index τ_n is chosen based on the data-driven procedure discussed in Section 3.3, using 500 subsamples of size 500.

Table 7: Black-white wage gap with age restriction and additional premarket factors

Without controlling for premarket factors, our estimation results show that the black-white wage gap has decreased by 4.6 points across the 1979 and 1997 cohorts. This result provides evidence of a very slow black-white wage convergence between 1990 and 2007. While most of the available evidence in the literature relates to the evolution of the black-white wage gap before 2000, it is interesting to note that our results are of the same order of magnitude as the estimates obtained by Neal & Rick (2014) using different datasets (namely the Census Long Form for the year 1990 and the American Community Survey for the year 2007). In their paper, Neal & Rick address the issue of differential selection into the workforce by examining the sensitivity of the median black-white wage gap to various imputation rules, which vary based on the fraction of (missing) wages supposed to fall below the median of the potential wage distribution. This type of sensitivity analysis cannot be used after adding controls for premarket factors, since in that case knowing the fraction of wages falling below or above the median is not sufficient to estimate the median wage gap.

While we find that the black-white wage gap increases over time after controlling for AFQT, Table 7 shows that the direction of the change is overturned when including other premarket factors in addition to the AFQT. Using our estimation method, the black-white wage gap is found to be fairly stable across cohorts, declining by only 1.2 points (from 9.9% to 8.7%) between 1990 and 2007. This result suggests that the across-cohort increase in the wage gap conditional on AFQT is actually attributable to the premarket factors other than AFQT, thus reflecting a time-varying omitted variable bias based on these family environment characteristics. Interestingly, one can understand this result as extending the key finding of Neal & Johnson (1996) to the across-cohort change in the wage gap. Premarket factors are a dominant component of the black-white wage gap, not only in level but also in evolution.

In sum, our estimation results provide evidence of (i) a slow convergence in the raw male black-white wage gap between 1990 and 2007 (Specification without premarket factors), and (ii) an even slower convergence in the residual portion of the black-white wage gap, which remains after controlling for premarket factors (Preferred specification). While we do find that differences in premarket factors are a key component of the black-white wage gap and, as such, should be a major focus from a policy standpoint, the fact that its residual portion remains virtually stable after almost 20 years is also concerning.

We conclude this section by examining whether one could alternatively estimate the across-cohort evolution of the black-white wage gap by applying the inverse density weighting scheme of Lewbel (2007), treating AFQT as a special regressor. Note that, in this context, AFQT appears to be the only potential candidate as a special regressor, thus ruling out the possible use of the special regressor method in the absence of controls for premarket skills. The large support condition would require the employment probability to be arbitrarily small for some values of the AFQT. Although there is some variation, we found that the conditional employment probability, estimated via nonparametric regression, remains very far from 0, specifically above 0.63 for both NLSY cohorts. This clearly indicates that this method could not be used in this context.

6 Concluding remarks

In this paper, we develop a new semiparametric inference method for location-scale models in the presence of sample selection. A key feature of our method is that it can be used in situations where one does not have access to an instrument for selection, nor to a large support regressor. Instead, the main identifying condition is based on selection being independent of the covariates for large values of the outcome. We show that this condition is typically mild

provided that selection is endogenous. Building on this identification strategy, we propose a simple estimation procedure, which combines quantile regressions in the tails, or extremal quantile regressions, with minimum distance. We establish the consistency and asymptotic normality of our estimators by extending the analysis of Chernozhukov (2005) to a setting with sample selection. The choice of an appropriate quantile index is important in this context, and we derive a data-driven procedure for this purpose. Importantly for the practical usefulness of our method, we show that our estimation procedure performs well even with fairly small samples.

Finally, we apply our method to the estimation of the black-white wage gap among males from the NLSY79 and NLSY97 cohorts. Correcting for selection into the workforce is key in this context since black males are more likely to dropout from the labor market than white males, and this difference has increased over time. Our estimation results show that premarket factors play a major role in explaining the magnitude of the black-white wage gap, as well as its evolution over time.

References

Ahn, H., & Powell, J. L. 1993. Semiparametric Estimation of Censored Selection Models With a Nonparametric Selection Mechanism. *Journal of Econometrics*, **58**(1-2), 3–29.

Altonji, J.G., & Blank, R.M. 1999. Race and Gender in the Labor Market. *In:* Ashenfelter, O., & Card, D. (eds), *Handbook of Labor Economics*, vol. 3. Elsevier.

Altonji, J.G., Ichimura, H., & Otsu, T. 2008. *Nonparametric Intermediate Order Regression Quantiles*. Working paper.

Altonji, J.G., Bharadwaj, P., & Lange, F. 2012. Changes in the Characteristics of American Youth: Implications for Adult Outcomes. *Journal of Labor Economics*, **30**, 783–828.

Andrews, D. K., & Schafgans, M. 1998. Semiparametric Estimation of the Intercept of a Sample Selection Model. *Review of Economic Studies*, **65**, 497–517.

Andrews, D.W.K. 1999. Consistent Moment Selection Procedures for Generalized Method of Moments Estimation. *Econometrica*, **67**(3), 543–564.

Andrews, D.W.K., & Soares, G. 2010. Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *Econometrica*, **78**(1), 119–157.

Arellano, M., & Bonhomme, S. 2011. *Quantile Selection Models*. Working paper.

Beaudry, P., Green, D.A., & Sand, B.M. 2013. The Great Reversal in the Demand for Skill and Cognitive Tasks. *NBER Working Paper No. 18901*.

Brown, C. 1984. Black-White Earnings Ratios Since the Civil Rights Act of 1964: The Importance of Labor Market Dropouts. *Quarterly Journal of Economics*, **99**, 31–44.

Buchinsky, M. 1998. The Dynamics of Changes in The Female Wage Distribution in the USA: a Quantile Regression Approach. *Journal of Applied Econometrics*, **13**, 1–30.

Butler, R., & Heckman, J.J. 1977. The Government's Impact on the Labor Market Status of Black Americans: A Critical Review. *In:* et al., F. E. Bloch (ed), *Equal Rights and Industrial Relations*. Madison, Wis.: Indus. Relations Res. Assoc.

Carneiro, P., Heckman, J.J., & Masterov, D.V. 2005. Labor Market Discrimination and Racial Differences in Premarket Factors. *Journal of Law and Economics*, **48**(1), 1–39.

Carneiro, P.M., Hansen, K.T., & Heckman, J.J. 2003. Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review*, **44**, 361–422.

Castex, G., & Dechter, E. 2014. The Changing Roles of Education and Ability in Wage Determination. *Forthcoming in the Journal of Labor Economics*, **32**(4), –.

Chamberlain, G. 1986. Asymptotic Efficiency in Semiparametric Model with Censoring. *Journal of Econometrics*, **32**, 189–218.

Chernozhukov, V. 2005. Extremal Quantile Regression. *The Annals of Statistics*, **33**(2), pp. 806–839.

Chernozhukov, V., & Du, S. 2008. Extremal quantiles and value-at-risk. *In:* Durlauf, S., & Blume, L. (eds), *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan Press.

Chernozhukov, V., & Fernandez-Val, I. 2011. Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks. *Review of Economic Studies*, **78**, 559–589.

Chernozhukov, V., S.Lee, & Rosen, A.M. 2013. Intersection Bounds: Estimation and Inference. *Econometrica*, **81**(2), 667–737.

Cunha, F., & Heckman, J.J. 2007. Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling. *Labour Economics*, **14**, 870–893.

Danielsson, J., de Haan, L., Peng, L., & de Vries, C. G. 2001. Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis*, **76**, 226–248.

Das, M., Newey, W., & Vella, F. 2003. Nonparametric Estimation of Sample Selection Models. *Review of Economic Studies*, **70**, 33–58.

D'Haultfoeuille, X., & Maurel, A. 2013a. Another Look at the Identification at Infinity of Sample Selection Models. *Econometric Theory*, **29**(1), 213–224.

D'Haultfoeuille, X., & Maurel, A. 2013b. Inference on an Extended Roy Model, with an Application to Schooling Decisions in France. *Journal of Econometrics*, **174**, 95–106.

Drees, H., & Kaufmann, E. 1998. Selection of the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, **75**, 149–172.

Eisenhauer, P., Heckman, J.J., & Vytlacil, E.J. 2013. *The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs*. Working paper.

Gronau, R. 1974. Wage Comparison - a Selectivity Bias. *Journal of Political Economy*, **82**, 1119–43.

Heckman, J. J. 1974. Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, **42**, 679–694.

Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, **47**, 153–161.

Heckman, J. J. 1990. Varieties of Selection Bias. *The American Economic Review*, **80**, 313–318.

Heckman, J.J., & Vytlacil, E.J. 2007. Econometric evaluation of social programs, Part II. *In:* Heckman, J.J., & Leamer, E.E. (eds), *Handbook of Econometrics*, vol. 6B. Elsevier.

Johnson, W., Kitamura, Y., & Neal, D. 2000. Evaluating a Simple Method for Estimating Black-White Gaps in Median Wages. *American Economic Review*, **90**, 339–343.

Juhn, C. 2003. Labor Market Dropouts and Trends in the Wages of Black and White Men. *Industrial and Labor Relations Review*, **56**(4), 643–662.

Khan, S., & Tamer, E. 2010. Irregular Identification, Support Conditions and Inverse Weight Estimation. *Econometrica*, **78**, 2021–2042.

Kitagawa, T. 2010. *Testing for Instrument Independence in the Selection Model*. Working paper.

Lewbel, A. 2007. Endogenous Selection or Treatment Model Estimation. *Journal of Econometrics*, **141**, 777–806.

Lewbel, A. 2014. An Overview of the Special Regressor Method. *In:* Racine, J., Su, L., & Ullah, A. (eds), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press.

Neal, D. 2004. The Measured Black-White Wage Gap among Women Is Too Small. *Journal of Political Economy*, **112**, S1–S28.

Neal, D. 2006. Why Has Black-White Skill Convergence Stopped? *In:* Hanushek, E., & Welch, F. (eds), *Handbook of the Economics of Education*, vol. 1. Elsevier.

Neal, D., & Johnson, W.R. 1996. The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, **104**(5), pp. 869–895.

Neal, D., & Rick, A. 2014. *The Prison Boom and the Lack of Black Progress since Smith and Welch*. Working paper.

Nelsen, Roger B. 2006. *An Introduction to Copulas*. 2nd ed. edn. New York: Springer,.

Pollard, D. 1991. Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory*, **7**(2), pp. 186–199.

Resnick, S. 1987. *Extreme values, regular variation, and point processes*. New York: Springer-Verlag.

Schafgans, M., & Zinde-Walsh, V. 2002. On intercept estimation in the sample selection model. *Econometric Theory*, **18**, 40–50.

Smith, J., & Welch, F. 1989. Black Economic Progress After Myrdal. *Journal of Economic Literature*, **27**(2), 519–564.

Vella, F. 1998. Estimating Models with Sample Selection Bias: a Survey. *Journal of Human Resources*, **33**, 127–169.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

A Additional details on inference

A.1 Confidence intervals on quantile effects when $\delta_j \neq 0$

The endpoints of the confidence intervals on $\Delta_{j\tau}$ can be estimated by applying the methodology of Chernozhukov *et al.* (2013) to the outer bounds (2.3). We illustrate the procedure by focusing on the upper bound of the confidence interval, when $\delta_j > 0$. The lower bound of the interval and the case $\delta_j < 0$ can be treated similarly. Let

$$\theta_j(x) = \beta_j + \delta_j \frac{Q_{Y|D=1,X=x}(\frac{\tau}{P(D=1|X=x)}) - x'\beta}{1 + x'\delta},$$

so that $\overline{\Delta}^o_{j\tau} = \inf_{x \in \text{Supp}(X)} \theta_j(x)$. One can estimate $\theta_j(.)$ by

$$\widehat{\theta}_j(x) = \widehat{\beta}_j + \widehat{\delta}_j \frac{\widehat{Q}_{Y|D=1,X=x}(\frac{\tau}{\widehat{P}(D=1|X=x)}) - x'\widehat{\beta}}{1 + x'\widehat{\delta}},$$

where $\widehat{Q}_{Y|D=1,X=x}$ and $\widehat{P}(D=1|X=x)$ are nonparametric (for instance kernel) estimators of $Q_{Y|D=1,X=x}$ and $P(D=1|X=x)$, respectively. By Theorem 3.1, the rates of convergence of $\widehat{\delta}$ and $\widehat{\beta}$ are $(\tau_n n)^{-1/2}$ and $\gamma(\tau_n)(\tau_n n)^{-1/2}$, respectively. The rates of convergence of $\widehat{Q}_{Y|D=1,X=x}$ and $\widehat{P}(D=1|X=x)$ depend on the number of continuous components of X, on the degree of smoothness of $x \mapsto Q_{Y|D=1,X=x}$ and $x \mapsto P(D=1|X=x)$ as well as on the choice of the tuning parameters. In any case, it is always possible to choose τ_n so that the rate of convergence of $\widehat{\beta}$ will be slower than the rates of convergence of $\widehat{\delta}$, $\widehat{Q}_{Y|D=1,X=x}$ and $\widehat{P}(D=1|X=x)$.[16] In this case,

$$\theta_j(x) - \widehat{\theta}_j(x) = f_j(x)(\widehat{\beta} - \beta) + o_P\left(\frac{\gamma(\tau_n)}{\sqrt{\tau_n n}}\right),$$

where $f_j(x) = -e_j' + \frac{\widehat{\delta}_j x'}{1+x'\widehat{\delta}}$ and e_j is a vector of $\mathbb{R}^d$ whose jth coordinator equals 1 and others equal 0. One can then apply the inference procedure discussed in Section 4.1 of Chernozhukov *et al.* (2013) to construct the upper bound of the confidence interval. Note that although the rate of convergence is not $\sqrt{n}$ here, their Theorem 4.1 still applies, after replacing $\sqrt{n}$ by

[16] This is the case for any τ_n satisfying Assumption 9 if all the components of X are discrete. If one component of X is continuous, the rate of convergence for $Q_{Y|D=1,X=x}$ and $P(D=1|X=x)$ will typically be $n^{-2/5}$. Then one has to impose, in addition to Assumption 9, that $\tau_n = o(n^{-1/5})$. Under these conditions, one can show that such a τ_n always exist by a simple monotonicity argument.

$\sqrt{\tau_n n}/\gamma(\tau_n)$.[17] Specifically, let us define

$$s_n(x) = \frac{\gamma(\tau_n)}{\sqrt{\tau_n n}} \left\| f_j(x)\widehat{\Omega}_\delta^{1/2} \right\|_2, \quad Z_n^\star(x) = \frac{f_j(x)\widehat{\Omega}_\delta^{1/2}}{\left\| f_j(x)\widehat{\Omega}_\delta^{1/2} \right\|_2}\mathcal{N}_d.$$

where $\|.\|_2$ denotes the Euclidean norm, $\widehat{\Omega}_\delta$ is the consistent estimator of the asymptotic variance matrix Ω_δ described after Theorem 3.1 and $\mathcal{N}_d$ is a d-dimensional standard normal vector generated independently from the data. Then one can compute, typically by simulations,

$$K_{1n} = Q_{\sup_{x\in \text{Supp}(X)} Z_n^\star(x)|\text{data}}(1 - 0.1/\log(n)).$$

Now, constructing $\widehat{\mathcal{X}}_n$ as

$$\widehat{\mathcal{X}}_n = \left\{ x \in \text{Supp}(X) : \widehat{\theta}_j(x) \leq 2K_{1n}s_n(x) + \inf_{\widetilde{x}\in \text{Supp}(X)} \widehat{\theta}_j(\widetilde{x}) + K_{1n}s_n(\widetilde{x}) \right\}.$$

one can compute

$$K_{2n}(\tau) = Q_{\sup_{x\in\widehat{\mathcal{X}}_n} Z_n^\star(x)|\text{data}}(\tau).$$

Finally, the upper bound $\widehat{\overline{\theta}}_{j,1-\alpha}$ of a confidence interval on $\Delta_{j\tau}$ of nominal coverage $1-\alpha$ is defined by

$$\widehat{\overline{\theta}}_{j,1-\alpha} = \inf_{x\in\text{Supp}(X)} \widehat{\theta}_j(x) + K_{2n}(1-\alpha)s_n(x).$$

A.2 Details on the data-driven τ_n

We provide in this section a rationale for the construction of the data-driven τ_n detailed in Section 3.3. We study for that purpose the asymptotic behavior of $\widehat{\delta}$ for sequences τ_n' that do not satisfy Assumption 9 (iii), but only $\sqrt{\tau_n' n} f(\gamma(\tau_n')) = O(1)$. We show that in this case, $\widehat{\delta}$ has an asymptotic bias. Then we relate this bias with the asymptotic behavior of the J test statistic $T_J(\tau_n')$, and show how this can be used to select a quantile index for which the asymptotic bias is small.

First, let us define

$$\mu(\tau) \equiv \frac{E\left[(\tau - \mathbb{1}\{-Y \leq \gamma(\tau) + X'\beta(\tau))\,\overline{X}\right]}{\tau} = \frac{E\left[(\tau - \mathbb{1}\{-D\varepsilon \leq Q_{\widetilde{\varepsilon}}(\tau/h)(1+X'\delta))\,\overline{X}\right]}{\tau}.$$

[17] For that purpose, we need to assume their Condition V, which is a mild regularity condition (see Chernozhukov *et al.* , 2013, p.691, for a discussion). Then, using the proof of Theorem 3.1, we can check that under Assumptions 1-4 and 7-9, their Conditions P-(ii)-(v) hold (replacing $\sqrt{n}$ by $\sqrt{\tau_n n}/\gamma(\tau_n)$). Although Condition P-(i) does not hold, we can still prove their Lemma 4 using the fact that the nonparametric part of $\theta(.)$ does not play any role in the asymptotic distribution of $\widehat{\theta}(.)$.

As shown in the proof of Theorem 3.1, $\mu(\tau)$ is the core component of the bias induced by the fact that (3.2) is an equivalence instead of an equality. $f(\gamma(\tau_n))$ in Assumption 9 can then be viewed as an envelope of $\mu(\tau)$. Under Assumption 9-(iii), $\sqrt{\tau_n n}\mu(\tau_n) \to 0$, meaning that the asymptotic bias vanishes. In what follows, we derive the asymptotic bias of our estimator $\widehat{\delta}$ as a function of $\mu(\tau)$ and propose a subsampling method to approximate this bias.

From (C.10) and the linear representation of $\widehat{Z}_n(1)$ below (C.13) in the proof of Theorem 3.1, we have, for any sequence τ_n that satisfies Assumption 9,

$$\sqrt{\tau_n n}(\widehat{\delta} - \delta) = \log(m)(G'W_\delta^* G)^{-1} G'W_\delta^* \alpha_n(\tau_n) g_n(\delta) + o_P(1),$$

where $\alpha_n(\tau) = \sqrt{\tau_n n}/(\gamma(m\tau) - \gamma(\tau))$. We also show in the proof of Theorem 3.1 that

$$\alpha_n(\tau_n) g_n(\delta) = (I \otimes \Delta)\Gamma \widehat{Z}_n(\tau_n),$$

where $\widehat{Z}_n(\tau_n)$ is asymptotically normal with mean 0 when the asymptotic bias of $\widehat{\delta}$ is zero. In order to analyze situations where a sequence τ_n' only satisfies $\sqrt{\tau_n' n} f(\gamma(\tau_n')) = O(1)$, consider

$$\tilde{Z}_n(\tau) \equiv \log(m)\widehat{Z}_n(\tau) + Q_H^{-1}\sqrt{\tau n} b(\tau),$$

with $b(\tau) = (\mu(\tau), \sqrt{l_1}\mu(l_1\tau), \cdots, \sqrt{l_J}\mu(l_J\tau))'$. Then one can show that

$$\begin{aligned}\sqrt{\tau_n' n}(\widehat{\delta} - \delta) =& (G'W_\delta^* G)^{-1} G'W_\delta^* (I_J \otimes \Delta)\Gamma \tilde{Z}_n(\tau_n') \\ &- (G'W_\delta^* G)^{-1} G'W_\delta^* (I_J \otimes \Delta)\Gamma (I_{J+1} \otimes Q_H^{-1})\sqrt{\tau_n' n} b(\tau_n') + o_P(1).\end{aligned}$$

$\tilde{Z}_n(\tau_n')$ is asymptotically normal with mean 0 by definition of $b(\tau)$. Hence, the second term is the asymptotic bias of $\widehat{\delta}$. We seek to approximate the norm of this bias. In order to do so, we consider the minimum distance statistic, which is commonly used to conduct a specification test in the context of minimum distance estimation. By plugging in the minimum distance estimator $\widehat{\delta}$, we obtain

$$\begin{aligned}\log(m)\alpha_n(\tau_n') W_\delta^{*1/2} g_n(\widehat{\delta}) =& (I_{Jd} - W_\delta^{*1/2} G(G'W_\delta^* G)^{-1} G'W_\delta^{*1/2}) \left[(I \otimes \Delta)\Gamma \tilde{Z}_n(\tau_n') - \sqrt{\tau_n' n} B(\tau_n')\right] \\ &+ o_P(1),\end{aligned}$$

where $B(\tau) = (I \otimes \Delta)\Gamma(I_{J+1} \otimes Q_H^{-1})b(\tau)$. $B(\tau)$ is the bias associated with the choice of quantile index τ. It follows that the J-statistic defined in Section 3.3 can be written as

$$
\begin{aligned}
T_J(\tau_n') =& \tilde{Z}_n(\tau_n')'\Gamma'(I \otimes \Delta')(I_{Jd} - W_\delta^{*1/2} G(G'W_\delta^* G)^{-1} G' W_\delta^{*1/2})(I \otimes \Delta)\Gamma \tilde{Z}_n(\tau_n') \\
&+ 2\tilde{Z}_n(\tau_n')'\Gamma'(I \otimes \Delta')(I_{Jd} - W_\delta^{*1/2} G(G'W_\delta^* G)^{-1} G' W_\delta^{*1/2})\sqrt{\tau_n' n} B(\tau_n') \\
&+ \tau_n' n B(\tau_n')'(I_{Jd} - W_\delta^{*1/2} G(G'W_\delta^* G)^{-1} G' W_\delta^{*1/2}) B(\tau_n') + o_P(1).
\end{aligned}
$$

This equation shows that the J-statistic on the left-hand side converges to a chi square distribution with $(J-1)d$ degrees of freedom, plus a bias term. If $\sqrt{\tau_n' n} B(\tau_n') \to 0$, then the median of the J-statistic is asymptotically the median $M_{(J-1)d}$ of a $\chi^2((J-1)d)$. On the other hand, if the asymptotic bias $\sqrt{\tau_n' n} B(\tau_n')$ does not vanish, the difference between the median of the J-statistic and $M_{(J-1)d}$ will generally be asymptotically different from zero. Following this idea, we estimate the difference between the two medians and use it as a proxy for the asymptotic bias of $\widehat{\delta}$. As indicated in the text, we rely for that purpose on subsampling.

B Data appendix

We construct our NLSY97 dataset based on the interviews that were conducted during the years 2007 and 2008, using data on males from the cross-sectional sample and the oversample of blacks and Hispanics of the NLSY97. Our sample consists of the respondents who reported wages for at least one of these two years, along with the respondents who reported not working in either year (nonparticipants). Respondents with a missing AFQT score are excluded from the analysis. For the individuals working in both years, the wage variable is defined as the average of the hourly wages corresponding to the main job at the time of the interview. For those working during one year only, we define the wage variable as the hourly wage corresponding to the main job at the time of the interview in that year. Finally, we trim the data by dropping the wage observations below 1 dollar and above 118.95 dollars (corresponding to 75 dollars in 1991). We report in Table 8 below some descriptives corresponding to our NLSY79 and NLSY97 samples restricted to the respondents who took the ASVAB test when they were 16 or 17. Table 9 reports the labor force participation rates for the NLSY79 and NLSY97 samples, separately for blacks and whites.

	NLSY79			NLSY97		
	Blacks	Hispanics	Whites	Blacks	Hispanics	Whites
AFQT	-0.716	-0.314	0.387	-0.726	-0.279	0.373
Std.dev.	(0.812)	(0.935)	(0.966)	(1.037)	(0.942)	(0.923)
Highest grade completed	11.638	11.419	12.859	11.239	11.502	12.849
Std.dev.	(3.927)	(3.953)	(3.691)	(4.811)	(4.529)	(4.726)
Mother high school graduate	0.447	0.243	0.715	0.707	0.536	0.829
Father high school graduate	0.368	0.284	0.665	0.518	0.396	0.758
Mother college graduate	0.040	0.027	0.093	0.107	0.094	0.217
Father college graduate	0.046	0.050	0.188	0.086	0.068	0.245
Both parents at age 14	0.486	0.599	0.760	0.264	0.540	0.597

Note: Samples restricted to males. Blacks account for 31% (25%) of the NLSY79 (NLSY97) sample, while Hispanics account for 20% (21%) of the NLSY79 (NLSY97) samples.

Table 8: Descriptive statistics for the subsample with restricted age

	Blacks	Whites
NLSY79 full sample	91.02%	97.52%
NLSY79 with age restriction	90.58%	98.10%
NLSY97 with age restriction	81.43%	93.09%

Table 9: Labor force participation rates (males)

C Proofs of the results

In the following, we let, for any random variable U and with a slight abuse of notations, $S_U^{\leftarrow} = 1 - F_U^{\leftarrow}$. We also let $\tilde{U} = -U$ and define $\breve{\varepsilon} = \tilde{Y} + X'\beta$. Finally, we take the convention that intervals $[a, b]$ refer to $[b, a]$ when $b < a$, and similarly for open or semi-open intervals.

C.1 Proof of Theorem 2.1

By Theorem 2.1 in D'Haultfoeuille & Maurel (2013a), $x \mapsto x'\beta$ and $x \mapsto x'\delta$ are identified.[18] Identification of β and δ then follows from Assumption 2. Turning to $\Delta_{j\tau}$, remark first that by independence between X and ε,

$$\Delta_{j\tau} = \beta_j + \delta_j Q_\varepsilon(\tau). \tag{C.1}$$

[18] In D'Haultfoeuille & Maurel (2013a) we use $E(\exp(\lambda\varepsilon)) < +\infty$ for some $\lambda > 0$ instead of the weaker condition that $S_{\exp(\varepsilon)}$ slowly varying. An inspection of the proof reveals however that the proof only relies on this latter condition.

It suffices therefore to obtain bounds on $Q_\varepsilon(\tau)$. We suppose hereafter that $\delta_j > 0$. The reasoning is similar for $\delta_j < 0$, while Δ_j is identified from (C.1) if $\delta_j = 0$. First, we have

$$f_{\varepsilon|X}(u|x) = P(D=1|X=x) f_{\varepsilon|D=1,X}(u|x) + P(D=0|X=x) f_{\varepsilon|D=0,X}(u|x).$$

Thus, for all x in the support of X,

$$f_{\varepsilon|X}(u|x) \geq P(D=1|X=x) f_{\varepsilon|D=1,X}(u|x).$$

By independence between X and ε,

$$f_\varepsilon(u) = \sup_{x \in \text{Supp}(X)} f_{\varepsilon|X}(u|x) \geq \sup_{x \in \text{Supp}(X)} P(D=1|X=x) f_{\varepsilon|D=1,X}(u|x). \tag{C.2}$$

Integrating (C.2) between $-\infty$ and v implies that $F_\varepsilon(v) \geq \underline{F}_\varepsilon(v)$. Hence, $Q_\varepsilon(\tau) \leq \underline{F}_\varepsilon^{\leftarrow}(\tau)$. This yields the upper bound on $\Delta_{j\tau}$. Now, integrating (C.2) between v and $+\infty$ implies that

$$1 - F_\varepsilon(v) \geq \int_v^{+\infty} \left[\sup_{x \in \text{Supp}(X)} P(D=1|X=x) f_{\varepsilon|D=1,X}(u|x) \right] du.$$

Hence, $F_\varepsilon(v) \leq 1 - \underline{F}_\varepsilon(+\infty) + \underline{F}_\varepsilon(v)$. As a result,

$$Q_\varepsilon(\tau) \geq \left[1 - \underline{F}_\varepsilon(+\infty) + \underline{F}_\varepsilon\right]^{\leftarrow}(\tau) = \underline{F}_\varepsilon^{\leftarrow}(\tau - (1 - \underline{F}_\varepsilon(+\infty))).$$

The lower bound on $\Delta_{j\tau}$ follows.

Now, let us show that these bounds are sharp under (2.2). For that purpose, we exhibit conditional cdfs $\widetilde{F}_{\varepsilon|D=0,X}(.|.)$, different in general from the true ones, which rationalize the bounds and satisfy the restrictions imposed by Assumptions 1, 3, 4 and (2.2). Note that the other conditions (Assumptions 2 and 3-(iv)) only depend on the observed data and therefore need not be verified. Note also that we can restrict to the case where $P(D=0|X=x) > 0$ for almost all x. For in the case where $P(D=0|X=x) = 0$, Inequality (C.2) is actually an equality, and the two bounds coincide. The bounds then correspond to the true model, and are therefore sharp.

Now, consider the upper bound. Let u_0 be such that $\underline{F}_\varepsilon(u_0) > \tau$ and suppose that

$$\widetilde{F}_{\varepsilon|D=0,X}(u|x) = \frac{\underline{F}_\varepsilon(u) - P(\varepsilon \leq u, D=1|X=x)}{P(D=0|X=x)} \mathbb{1}\{u < u_0\} + F_{\varepsilon|D=0,X}(u|x) \mathbb{1}\{u \geq u_0\}.$$

Let us first show that for all x, $\widetilde{F}_{\varepsilon|D=0,X}(.|x)$ is indeed a cdf. It suffices to show that its limit at

$-\infty$ is zero, that it is increasing and right-continuous on $(-\infty, u_0)$ and $\lim_{u\uparrow u_0} \widetilde{F}_{\varepsilon|D=0,X}(u|x) \leq F_{\varepsilon|D=0,X}(u_0|x)$. The first point holds because $\lim_{u\to-\infty} \underline{F}_\varepsilon(u) = \lim_{u\to-\infty} P(\varepsilon \leq u, D = 1|X = x) = 0$. The second point follows by remarking that

$$\underline{F}_\varepsilon(v) - P(\varepsilon \leq v, D = 1|X = x) = \int_{-\infty}^{v} \Bigg\{ \sup_{x' \in \text{Supp}(X)} \left[f_{\varepsilon|D=1,X}(u|x') P(D = 1|X = x') \right] - f_{\varepsilon|D=1,X}(u|x) P(D = 1|X = x) \Bigg\} du.$$

The integral form implies that $\widetilde{F}_{\varepsilon|D=0,X}(.|x)$ is right-continuous. Because the term in braces is positive, $\widetilde{F}_{\varepsilon|D=0,X}(.|x)$ is also increasing. Finally, the third point follows because for any u,

$$\frac{\underline{F}_\varepsilon(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)} \leq \frac{F_\varepsilon(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)} = F_{\varepsilon|D=0,X}(u|x).$$

Now, let us prove that the conditional cdfs $\widetilde{F}_{\varepsilon|D=0,X}(.|.)$ rationalize the bounds and satisfy the restrictions of the model. First,

$$\begin{aligned} \widetilde{F}_{\varepsilon|X}(u|x) &= \widetilde{F}_{\varepsilon|D=0,X}(u|x) P(D = 0|X = x) + P(\varepsilon \leq u, D = 1|X = x) \\ &= \underline{F}_\varepsilon(u) \mathbb{1}\{u < u_0\} + F_\varepsilon(u) \mathbb{1}\{u \geq u_0\}. \end{aligned} \tag{C.3}$$

The right-hand side does not depend on x. Therefore, $\widetilde{F}_{\varepsilon|D=0,X}$ satisfies Assumption 1. (C.3) also implies that for any $\tau' \leq \tau$,

$$\widetilde{F}_\varepsilon^{\leftarrow}(\tau') = \underline{F}_\varepsilon^{\leftarrow}(\tau').$$

Therefore, the conditional cdfs $\widetilde{F}_{\varepsilon|D=0,X}(.|.)$ rationalize $\overline{\Delta}_j(\tau)$. Now, because $\widetilde{f}_\varepsilon(u)$ is equal to the true $f_\varepsilon(u)$ for u large enough, the conditional cdfs $\widetilde{F}_{\varepsilon|D=0,X}(.|.)$ satisfy Assumption 3. Similarly, by Bayes' theorem, we have, for y large enough,

$$\begin{aligned} \widetilde{P}(D = 1|X = x, Y^* = y) &= \frac{f_{Y|D=1,X=x}(y|x) P(D = 1|X = x)}{\widetilde{f}_{Y^*|X}(y|x)} \\ &= \frac{(1 + x'\delta) f_{Y|D=1,X=x}(y|x) P(D = 1|X = x)}{\widetilde{f}_\varepsilon[(y - x'\beta)/(1 + x'\delta)]} \\ &= \frac{(1 + x'\delta) f_{Y|D=1,X=x}(y|x) P(D = 1|X = x)}{f_\varepsilon[(y - x'\beta)/(1 + x'\delta)]} \\ &= P(D = 1|X = x, Y^* = y), \end{aligned}$$

and therefore, Assumption 4 is satisfied. This equality also ensures that (2.2) is satisfied. Hence, the upper bound is sharp.

Now, let us turn to the lower bound. Let u_1 be such that $\underline{F}_\varepsilon(u_1) < \tau$ and consider

$$\widetilde{F}_{\varepsilon|D=0,X}(u|x) = F_{\varepsilon|D=0,X}(u|x)\mathbb{1}\{u < u_1\} + \frac{1 - \underline{F}_\varepsilon(+\infty) + \underline{F}_\varepsilon(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)}\mathbb{1}\{u \geq u_1\}.$$

As previously, $\widetilde{F}_{\varepsilon|D=0,X}(.|x)$ is indeed a cdf and

$$\widetilde{F}_{\varepsilon|X}(u|x) = F_\varepsilon(u)\mathbb{1}\{u < u_1\} + \left[1 - \underline{F}_\varepsilon(+\infty) + \underline{F}_\varepsilon(u)\right]\mathbb{1}\{u \geq u_1\},$$

so that Assumption 1 holds and $\widetilde{F}_{\varepsilon|D=0,X}(.|.)$ rationalizes the lower bound. We now check Assumption 3. For u large enough,

$$\begin{aligned}
\widetilde{f}_\varepsilon(u) &= \sup_{x \in \text{Supp}(X)} P(D = 1|X = x) f_{\varepsilon|D=1,X}(u|x) \\
&= \sup_{x \in \text{Supp}(X)} P(D = 1|X = x, \varepsilon = u) f_{\varepsilon|X}(u|x) \\
&= f_\varepsilon(u) \left[\sup_{x \in \text{Supp}(X)} P(D = 1|X = x, Y^* = x'\beta + (1 + x'\delta)u) \right]. \qquad \text{(C.4)}
\end{aligned}$$

We now prove that $\widetilde{f}_\varepsilon(u) \sim h f_\varepsilon(u)$ as $u \to \infty$. Fix $\eta > 0$. Because $\text{Supp}(X)$ is compact, there exists $(x_1, ..., x_k) \in \text{Supp}(X)^k$ such that for all $x \in \text{Supp}(X)$, $\min_{j=1...k} \|x - x_j\| < \eta$. There exists also y_0 such that for all $y \geq y_0$,

$$\max_{j=1...k} |P(D = 1|X = x_j, Y^* = y) - h| < \eta. \qquad \text{(C.5)}$$

By compacity of $\text{Supp}(X)$ once more, there exists u_0 such that for all $u \geq u_0$,

$$\inf_{x \in \text{Supp}(X)} x'\beta + (1 + x'\delta)u \geq y_0. \qquad \text{(C.6)}$$

Then, for all $x \in \text{Supp}(X)$, and all $u \geq u_0$,

$$\begin{aligned}
&|P(D = 1|X = x, Y^* = x'\beta + (1 + x'\delta)u) - h| \\
&\leq |P(D = 1|X = x, Y^* = x'\beta + (1 + x'\delta)u) - P(D = 1|X = x_j, Y^* = x'\beta + (1 + x'\delta)u)| \\
&\quad + |P(D = 1|X = x_j, Y^* = x'\beta + (1 + x'\delta)u) - h| \\
&\leq K\|x - x_j\| + \eta,
\end{aligned}$$

where the second inequality follows by (2.2), (C.5) and (C.6). Choosing j such that $\|x - x_j\| <$

η finally yields

$$\sup_{x\in\text{Supp}(X)} |P(D=1|X=x, Y^*=x'\beta+(1+x'\delta)u)-h| < (K+1)\eta.$$

As a result,

$$\lim_{u\to\infty} \sup_{x\in\text{Supp}(X)} P(D=1|X=x, Y^*=x'\beta+(1+x'\delta)u) = h.$$

Hence, by (C.4), as $u\to\infty$,

$$\widetilde{f}_\varepsilon(u) \sim h f_\varepsilon(u). \tag{C.7}$$

This implies that Assumption 3-(i) holds. Now, suppose that $\widetilde{S}_{\exp(\varepsilon)}$ is slowly varying. Then for all $l>0$, $\widetilde{S}_{\exp(\varepsilon)}(lu)/\widetilde{S}_{\exp(\varepsilon)}(u)\to 1$. Now, (C.7) also implies that for any $l>0$,

$$\frac{S_{\exp(\varepsilon)}(lu)}{S_{\exp(\varepsilon)}(u)} \sim \frac{\widetilde{S}_{\exp(\varepsilon)}(lu)}{\widetilde{S}_{\exp(\varepsilon)}(u)}.$$

This implies that $S_{\exp(\varepsilon)}$ is also slowly varying, a contradiction. Thus, Assumption 3-(ii) is satisfied. By (C.7) once more, there exists $\eta>0$ arbitrarily small such that for all u large enough,

$$(h-\eta)S_\varepsilon(u) \le \widetilde{S}_\varepsilon(u) \le (h+\eta)S_\varepsilon(u).$$

Now, fix τ small enough and let $u=\widetilde{S}_\varepsilon^{\leftarrow}(\tau)$. $\widetilde{S}_\varepsilon(u)\ge\tau$ implies $S_\varepsilon(u)\ge\tau/(h+\eta)$, which yields in turn $u\le S_\varepsilon^{\leftarrow}(\tau/(h+\eta))$. Hence, we obtain

$$\widetilde{S}_\varepsilon^{\leftarrow}(\tau) \le S_\varepsilon^{\leftarrow}(\tau/(h+\eta)) = -Q_{\widetilde{\varepsilon}}(\tau/(h+\eta)).$$

Now, let $u'>u$, so that $\widetilde{S}_\varepsilon(u')\le\tau$. Then $u'\ge S_\varepsilon^{\leftarrow}(\tau/(h-\eta))$. Letting u' tend to u yields

$$\widetilde{S}_\varepsilon^{\leftarrow}(\tau) \ge S_\varepsilon^{\leftarrow}(\tau/(h-\eta)) = -Q_{\widetilde{\varepsilon}}(\tau/(h-\eta)).$$

As a result, for any fixed $m>1$ and letting $e=\exp(1)$,

$$\frac{\widetilde{S}_\varepsilon^{\leftarrow}(m\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)}{\widetilde{S}_\varepsilon^{\leftarrow}(e\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)} \le \frac{Q_{\widetilde{\varepsilon}}(m\tau/(h-\eta))-Q_{\widetilde{\varepsilon}}(\tau/(h+\eta))}{Q_{\widetilde{\varepsilon}}(e\tau/(h+\eta))-Q_{\widetilde{\varepsilon}}(\tau/(h-\eta))}.$$

By Lemma D.2 in Appendix D, the right-hand side converges to $\log(m(h+\eta)/(h-\eta))/\log(e(h-$

$\eta)/(h+\eta))$. Reasoning similarly on the lower bound, we obtain,

$$\log(m(h-\eta)/(h+\eta))/\log(e(h+\eta)/(h-\eta)) \leq \liminf_{\tau\to 0} \frac{\widetilde{S}_\varepsilon^{\leftarrow}(m\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)}{\widetilde{S}_\varepsilon^{\leftarrow}(e\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)}$$
$$\leq \limsup_{\tau\to 0} \frac{\widetilde{S}_\varepsilon^{\leftarrow}(m\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)}{\widetilde{S}_\varepsilon^{\leftarrow}(e\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)} \leq \log(m(h+\eta)/(h-\eta))/\log(e(h-\eta)/(h+\eta)).$$

Because η was arbitrary, we can make it tend to zero, thus obtaining

$$\lim_{\tau\to 0} \frac{\widetilde{S}_\varepsilon^{\leftarrow}(m\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)}{\widetilde{S}_\varepsilon^{\leftarrow}(e\tau)-\widetilde{S}_\varepsilon^{\leftarrow}(\tau)} = \log(m).$$

This proves (see Resnick, 1987, Proposition 0.10) that $\widetilde{S}_\varepsilon$ belongs to the domain of attraction of the Gumbel distribution. Hence, Assumption 3-(iii) holds.

Turning to Assumption 4, we reason as for the upper bound:

$$\begin{aligned}\widetilde{P}(D=1|X=x,Y^*=y) &= \frac{(1+x'\delta)f_{Y|D=1,X=x}(y|x)P(D=1|X=x)}{\widetilde{f}_\varepsilon[(y-x'\beta)/(1+x'\delta)]}\\ &\sim \frac{(1+x'\delta)f_{Y|D=1,X=x}(y|x)P(D=1|X=x)}{hf_\varepsilon[(y-x'\beta)/(1+x'\delta)]}\\ &\sim \frac{P(D=1|X=x,Y^*=y)}{h}.\end{aligned}$$

Therefore, the conditional cdfs $\widetilde{F}_{\varepsilon|D=0,X}$ satisfy Assumption 4, with a limit equal to 1 instead of h. The result follows.

Finally, let us check (2.2). We have by what precedes, for y large enough

$$\widetilde{P}(D=1|X=x,Y^*=y) = \frac{P(D=1|X=x,Y^*=y)}{\sup_{x'\in \mathrm{Supp}(X)} P(D=1|X=x',Y^*=y)}.$$

Moreover, we have proved that the denominator tends to h as $y\to\infty$. Therefore, because the true distribution satisfies (2.2), we have, for all $(x,x')\in \mathrm{Supp}(X)^2$ and all y large enough,

$$|\widetilde{P}(D=1|X=x',Y^*=y)-\widetilde{P}(D=1|X=x,Y^*=y)| \leq \frac{K}{h-\eta}\|x-x'\|,$$

for some $0<\eta<h$. This ensures that $\widetilde{F}_{\varepsilon|D=0,X}$ satisfies (2.2), and thus that the lower bound is sharp.

C.2 Derivation of the outer bounds (2.3)

We only consider the case where $\delta_j > 0$, the case $\delta_j < 0$ being similar. Note first that $P(D = 1|X) > 0$ almost surely, because $P(D = 1|X = x) = E\left[P(D = 1|X = x, Y^*)|X = x\right]$, and $P(D = 1|X = x, Y^* = y)$ is bounded from below by $h/2 > 0$ for y large enough. Now, consider the lower bound. We have for all (u, x), by independence between ε and X,

$$\begin{aligned} P(\varepsilon \leq u) &= P(\varepsilon \leq u|X = x) \\ &\leq P(\varepsilon \leq u, D = 1|X = x) + P(D = 0|X = x) \\ &\leq P(Y \leq x'\beta + (1 + x'\delta)u|D = 1, X = x)P(D = 1|X = x) + P(D = 0|X = x). \end{aligned}$$

Taking $u = Q_\varepsilon(\tau)$, using $F_\varepsilon(Q_\varepsilon(\tau)) \geq \tau$ and the definition of the quantiles of $Y|D = 1, X = x$, we obtain, for all x in the support of X,

$$x'\beta + (1 + x'\delta)Q_\varepsilon(\tau) \geq Q_{Y|D=1,X=x}\left(\frac{\tau - P(D = 0|X = x)}{P(D = 1|X = x)}\right).$$

As a result,

$$Q_\varepsilon(\tau) \geq \sup_{x \in \text{Supp}(X)} \frac{Q_{Y|D=1,X=x}\left(\frac{\tau - P(D=0|X=x)}{P(D=1|X=x)}\right) - x'\beta}{1 + x'\delta}.$$

The outer lower bound of $\Delta_{j\tau}$ follows from $\Delta_{j\tau} = \beta_j + \delta_j Q_\varepsilon(\tau)$.

Now let us turn to the outer upper bound. Reasoning as before, we have, for all x in the support of X and $u < Q_\varepsilon(\tau)$,

$$\tau \geq P(\varepsilon \leq u) \geq P(Y \leq x'\beta + (1 + x'\delta)u, D = 1|X = x)P(D = 1|X = x).$$

The definition of the quantiles of $Y|D = 1, X = x$ then yields

$$x'\beta + (1 + x'\delta)u \leq Q_{Y|D=1,X=x}\left(\frac{\tau}{P(D = 1|X = x)}\right).$$

Letting u tend to $Q_\varepsilon(\tau)$ and taking the infimum over x then yields

$$Q_\varepsilon(\tau) \leq \inf_{x \in \text{Supp}(X)} \frac{Q_{Y|D=1,X=x}\left(\frac{\tau}{P(D=1|X=x)}\right) - x'\beta}{1 + x'\delta}.$$

The upper bound follows.

C.3 Proof of Proposition 2.1

We verify Assumption 4 with $h = 1$. By Assumption 5 and because $f_C(\gamma) \to 0$, we have, as $y \to \infty$,

$$\begin{aligned}
|P(D=1|X=x,Y^*=y)-1| &= \left|P\left(F_\eta(\eta) \leq F_\eta(\phi(x))|F_{\widetilde{\varepsilon}}(\widetilde{\varepsilon}) = F_{\widetilde{\varepsilon}}\left(\frac{x'\beta - y}{1+x'\delta}\right)\right) - 1\right| \\
&= \left|\partial_1 C\left[F_{\widetilde{\varepsilon}}\left(\frac{x'\beta - y}{1+x'\delta}\right), F_\eta(\phi(x))\right] - 1\right| \\
&\leq \sup_{v\in[\underline{v},1]}\left|\partial_1 C\left[F_{\widetilde{\varepsilon}}\left(\frac{x'\beta - y}{1+x'\delta}\right), v\right] - 1\right| \\
&\longrightarrow 0.
\end{aligned}$$

C.4 Proof of Proposition 2.2

Let $D_s = \mathbb{1}\{D = s\}$ and $Y = \sum_{s=1}^J D_s Y_s^*$. By considering the dataset $(D_s, D_s Y, X)$, we are back to the binary model. Then Theorem 2.1 is directly applicable if one can verify Assumption 4. For a given x in the support of X, we have

$$\begin{aligned}
&1 - P(D_s = 1|X = x, Y_s^* = y) \\
=&P(y + G_s(x) + \lambda'_{s,2}\pi + \nu_{s,2} \leq max_{i\neq s}\left(x'\beta_i + (1+x'\delta_i)(\lambda'_{i,1}\pi + \nu_{i,1}) + G_i(x) + \lambda'_{i,2}\pi + \nu_{i,2}\right)) \\
\leq&P(\nu_s \geq y + G(x)),
\end{aligned}$$

where

$$\begin{aligned}
\nu_s &= \sup_{x\in \text{Supp}(X)} \max_{i\neq s}((1+x'\delta_i)\nu_{i,1} + \nu_{i,2} - \nu_{s,2}), \\
G(x) &= \inf_{p\in\text{Supp}(\pi)}(G_s(x) + \lambda'_{s,2}p - \max_{i\neq s}\left(x'\beta_i + (1+x'\delta_i)\lambda'_{i,1}p + G_i(x) + \lambda'_{i,2}p\right)).
\end{aligned}$$

By Assumption 6, $G(x) > -\infty$. Therefore, as $y \to \infty$, $P(\nu_s \geq y + G(x)) \to 0$. Thus, Assumption 4 holds (with $h = 1$).

C.5 Proof of Lemma 3.1

Let $U_x(y) \equiv 1/P(Y > y|X = x)$, $V_x(y) \equiv 1/hS_\varepsilon((y - x\beta)/(1 + x'\delta))$. Then from Equation (3.1), $U_x(y) \sim V_x(y)$. We want to show the equivalence $U_x^{\leftarrow}(\tau) \sim V_x^{\leftarrow}(\tau)$. For that purpose, we suppose that there exists $\varepsilon_0 > 0$ and a sequence $(y_m)_{m\in\mathbb{N}}$ tending to infinity such that

$$V_x^{\leftarrow}(y_m)/U_x^{\leftarrow}(y_m) \geq 1 + \varepsilon_0, \tag{C.8}$$

and shows that this leads to a contradiction. The reasoning is similar for the other inequality $(V_x^{\leftarrow}(y_m)/U_x^{\leftarrow}(y_m) \leq 1-\varepsilon_0)$.

First, by Lemma D.1 in Appendix D, S_ε is in the domain of attraction of Type I extreme value distribution. This implies that $V \equiv 1/S_\varepsilon$ is Γ-varying (see Resnick (1987) Proposition 0.10), i.e. $\lim_{z\to\infty} \frac{V(z+tf(z))}{V(z)} = e^t$ for some auxiliary function f. Define $f_x(y) = f\left[(y-x'\beta)/(1+x'\delta)\right] \times (1+x'\delta)$. Then

$$\frac{V_x(z+tf_x(z))}{V_x(z)} = \frac{V\left[\frac{z-x'\beta}{1+x'\delta} + tf\left(\frac{z-x'\beta}{1+x'\delta}\right)\right]}{V\left[\frac{z-x'\beta}{1+x'\delta}\right]} \to e^t$$

as $z \to \infty$. Thus $V_x(z)$ is Γ-varying with auxiliary function f_x. Furthermore, $U_x(z) \sim V_x(z)$ and $z + tf_x(z) \to \infty$, which implies

$$\frac{U_x(z+tf_x(z))}{U_x(z)} = \frac{U_x(z+tf_x(z))}{V_x(z+tf_x(z))}\frac{V_x(z)}{U_x(z)}\frac{V_x(z+tf_x(z))}{V_x(z)} \to e^t.$$

Hence, U_x is also Γ-varying with the same auxiliary function. f_x also satisfies (see Resnick, 1987, Ex. 0.4.3.10)

$$\lim_{z\to\infty} \frac{f_x(z)}{z} \to 0. \tag{C.9}$$

Combining (C.8) and (C.9), we obtain that for m large enough,

$$\frac{V_x^{\leftarrow}(y_m)}{U_x^{\leftarrow}(y_m)} \geq 1 + \varepsilon_0 \frac{f_x(U_x^{\leftarrow}(y_m))}{U_x^{\leftarrow}(y_m)}.$$

Now, because $y \sim V_x(V_x^{\leftarrow}(y))$ and $y \sim U_x(U_x^{\leftarrow}(y))$ (see Resnick, 1987, page 28), for any $\varepsilon_1 > 0$, there exists m large enough such that

$$\begin{aligned}
y_m(1+\varepsilon_1) &\geq V_x(V_x^{\leftarrow}(y_m)) \\
&\geq V_x(U_x^{\leftarrow}(y_m) + \varepsilon_0 f_x(U_x^{\leftarrow}(y_m))) \\
&\geq (1-\varepsilon_1) U_x(U_x^{\leftarrow}(y_m) + \varepsilon_0 f_x(U_x^{\leftarrow}(y_m))) \\
&= (1-\varepsilon_1)^2 e^{\varepsilon_0} U_x(U_x^{\leftarrow}(y_m)) \\
&\geq (1-\varepsilon_1)^3 e^{\varepsilon_0} y_m.
\end{aligned}$$

Therefore, $1 \geq \frac{(1-\varepsilon_1)^3}{1+\varepsilon_1} e^{\varepsilon_0}$. Letting ε_1 tend to zero leads to a contradiction.

C.6 Proof of Theorem 3.1

First let us introduce additional notations. For any τ, let $\theta(\tau) = (\gamma(\tau), \beta(\tau)')'$. Let us also define $\widehat{Z}_n(l) = \alpha_n(l)(\widehat{\theta}(\tau_n) - \theta(\tau_n))$, with

$$\alpha_n(l) = \frac{\sqrt{l\tau_n n}}{\gamma(ml\tau_n) - \gamma(l\tau_n)} = \frac{\sqrt{l\tau_n n}}{Q_{\widetilde{\varepsilon}}(ml\tau/h) - Q_{\widetilde{\varepsilon}}(l\tau/h)}$$

for some arbitrary fixed $m > 1$ and $\alpha_n \equiv \alpha_n(1)$. Let also

$$\widehat{Z}_n(l_1, \cdots, l_J) = \left(\widehat{Z}_n'(1), \widehat{Z}_n'(l_1), ..., \widehat{Z}_n'(l_J)\right)'.$$

Finally, let us define

$$G_n \equiv -\frac{\partial g_n(\delta)}{\partial \delta} = \begin{pmatrix} \widehat{\gamma}(l_1\tau_n) - \widehat{\gamma}(\tau_n) \\ \vdots \\ \widehat{\gamma}(l_J\tau_n) - \widehat{\gamma}(\tau_n) \end{pmatrix} \otimes I_d,$$

$\widetilde{G}_n = G_n/(\gamma(m\tau_n) - \gamma(\tau_n))$ and

$$\widehat{B}_n = \begin{pmatrix} \widehat{\beta}(l_1\tau_n) - \widehat{\beta}(\tau_n) \\ \vdots \\ \widehat{\beta}(l_J\tau_n) - \widehat{\beta}(\tau_n) \end{pmatrix}.$$

The main part of the proof is devoted to the asymptotic normality of $\widehat{\delta}$. The asymptotic normality of $\widehat{\beta}$ and the second part of the theorem follows quite easily.

The behavior of $\widehat{\delta}$ is related to $\widetilde{G}_n$ and $\widehat{Z}_n(l_1, \cdots, l_J)$. To see this, note that the first order condition of (3.3) writes

$$G_n' W_n G_n \widehat{\delta} = G_n' W_n \widehat{B}_n.$$

Remarking that $\delta = [G_n' W_n G_n]^{-1} G_n' W_n G_n \delta$ and $\widehat{B}_n - G_n\delta = g_n(\delta)$, we obtain

$$\sqrt{\tau_n n}(\widehat{\delta} - \delta) = \left[\widetilde{G}_n' W_n \widetilde{G}_n\right]^{-1} \widetilde{G}_n' W_n \left[\alpha_n g_n(\delta)\right].$$

Moreover, some algebra shows that

$$\alpha_n g_n(\delta) = (I_J \otimes \Delta)\Gamma \widehat{Z}_n(l_1, \cdots, l_J).$$

We thus obtain

$$\sqrt{\tau_n n}(\widehat{\delta}-\delta)=\left[\widetilde{G}_n' W_n \widetilde{G}_n\right]^{-1}\widetilde{G}_n' W_n\,(I_J\otimes\Delta)\,\Gamma\,\widehat{Z}_n(l_1,\cdots,l_J). \tag{C.10}$$

The first step of the proof shows that $\widehat{Z}_n(l_1,\cdots,l_J)$ is asymptotically normal. The proof of this part is related to the proof of Theorem 5.1 in Chernozhukov (2005), but we have to take into account that (3.2) is an equivalence, not an equality as in his framework. The second step establishes that $\widetilde{G}_n \xrightarrow{p} G/\log(m)$. Both steps, combined with (C.10), prove the asymptotic normality of $\widehat{\delta}$. We then show in the third step the main asymptotic result on $\widehat{\beta}$. Finally, Step 4 establishes the consistency of $\widehat{\beta}$ and the fact that the rate of convergence of $\widehat{\delta}$ and $\widehat{\beta}$ can be polynomial under some additional conditions on $f(.)$.

1. $\widehat{Z}_n(l_1,\cdots,l_J)\xrightarrow{d}\mathcal{N}(0,\log(m)^{-2}L\otimes\Omega_0)$.

We prove the result for $\widehat{Z}_n(1)$ only, the multivariate generalization being straightforward but notationally cumbersome. Similarly to Chernozhukov (2005), Equation (9.43), $\widehat{Z}_n(1)$ minimizes

$$\Psi_n(z,\tau_n)=W_n(\tau_n)'z+\Lambda_n(z,\tau_n),$$

with, for any τ,

$$W_n(\tau)=\frac{-1}{\sqrt{\tau n}}\sum_{i=1}^{n}(\tau-\mathbb{1}\{(\tilde{Y}_i-\gamma(\tau)-X_i'\beta(\tau)\leq 0)\})\overline{X}_i \tag{C.11}$$

and for any $z=(z_1,z_2')'\in\mathbb{R}\times\mathbb{R}^d$,

$$\Lambda_n(z,\tau)=\frac{\alpha_n}{\sqrt{\tau n}}\sum_{i=1}^{n}\int_0^{(z_1+X_i'z_2)/\alpha_n}\mathbb{1}\{\tilde{Y}_i-\gamma(\tau)-X_i'\beta(\tau)\leq s\}-\mathbb{1}\{\tilde{Y}_i-\gamma(\tau)-X_i'\beta(\tau)\leq 0\}ds. \tag{C.12}$$

$\Lambda_n(z,\tau_n)$ is convex in z because the integrands are increasing in s. Moreover, by Lemma D.4 in Appendix D, $\Lambda_n(z,\tau_n)\to\frac{1}{2}\log(m)z'\mathcal{Q}_H z$. We shall now prove that

$$W_n(\tau_n)\xrightarrow{d}\mathcal{N}(0,\mathcal{Q}_X). \tag{C.13}$$

By applying the convexity lemma and the same arguments as in the end of the proof of Theorem 1 in Pollard (1991), Condition (C.13) implies $\widehat{Z}_n(1)+\log(m)^{-1}\mathcal{Q}_H^{-1}W_n(\tau_n)=o_P(1)$ and thus $\widehat{Z}_n(1)\xrightarrow{d}\mathcal{N}\left(0,\log(m)^{-2}\Omega_0\right)$.

To establish (C.13), let $M_{n,i}(\tau) = \frac{-1}{\sqrt{\tau n}}(\tau - \mathbb{1}\{\tilde{Y}_i - \gamma(\tau) - X_i'\beta(\tau) \leq 0\})\overline{X}_i - \sqrt{\tau/n}\mu(\tau)$, with

$$\mu(\tau) \equiv \frac{E\left[\left(\tau - \mathbb{1}\{\tilde{Y} \leq \gamma(\tau) + X'\beta(\tau)\}\right)\overline{X}\right]}{\tau}.$$

Then

$$W_n(\tau) = \sum_{i=1}^{n} M_{n,i}(\tau) + \sqrt{n\tau}\mu(\tau). \tag{C.14}$$

By Lemma 9.6 of Chernozhukov (2005), we have

$$\sum_{i=1}^{n} M_{n,i}(\tau_n) \xrightarrow{d} \mathcal{N}(0, \mathcal{Q}_X). \tag{C.15}$$

Besides,

$$\begin{aligned}
\|\mu(\tau)\| &= \frac{1}{\tau}\|E\left[(hF_{\widetilde{\varepsilon}}(\gamma(\tau)) - P(D=1, \widetilde{\varepsilon} \leq \gamma(\tau)|X))\ \overline{X}\right]\| \\
&= \frac{1}{\tau}\left\|E\left\{\overline{X}\int_{-\infty}^{\gamma(\tau)}[h - P(D=1|X, \widetilde{\varepsilon}=e)]\, dF_{\widetilde{\varepsilon}}(e)\right\}\right\| \\
&\leq \frac{1}{\tau}E\left\{\|\overline{X}\| \sup_{e \leq \gamma(\tau)} |h - P(D=1|X, \widetilde{\varepsilon}=e)|\right\} F_{\widetilde{\varepsilon}}(\gamma(\tau)) \\
&= \frac{1}{h}f(\gamma(\tau)).
\end{aligned}$$

By Assumption 9, $\sqrt{n\tau_n}f(\gamma(\tau_n)) = o(1)$. Combined with (C.14) and (C.15), this proves (C.13).

2. $\widetilde{G}_n \xrightarrow{p} G/\log(m)$ and asymptotic normality of $\widehat{\delta}$.
First,

$$\frac{\widehat{\gamma}(l\tau_n) - \widehat{\gamma}(\tau_n)}{\gamma(m\tau_n) - \gamma(\tau_n)} = \frac{\widehat{\gamma}(l\tau_n) - \gamma(l\tau_n)}{\gamma(m\tau_n) - \gamma(\tau_n)} + \frac{\gamma(l\tau_n) - \gamma(\tau_n)}{\gamma(m\tau_n) - \gamma(\tau_n)} + \frac{\gamma(\tau_n) - \widehat{\gamma}(\tau_n)}{\gamma(m\tau_n) - \gamma(\tau_n)}. \tag{C.16}$$

Besides, by definition of α_n, $\widehat{\theta}(.)$ and $\widehat{Z}_n(l)$, Step 1 of the proof and because $\tau_n n \to \infty$ by Assumption 9,

$$\frac{\widehat{\gamma}(l\tau_n) - \gamma(l\tau_n)}{\gamma(m\tau_n) - \gamma(\tau_n)} = \frac{e_1'\alpha_n(\widehat{\theta}(l\tau_n) - \theta(l\tau_n))}{\sqrt{\tau_n n}} = \frac{e_1'\widehat{Z}_n(l)}{\tau_n n} = o_P(1),$$

where $e_1 = (1, 0, ..., 0)'$. Similarly, the third term of (C.16) also tends to zero in probability.

Now, by Lemma D.3 in Appendix D,

$$\frac{\gamma(l\tau_n)-\gamma(\tau_n)}{\gamma(m\tau_n)-\gamma(\tau_n)}=\frac{Q_{\widetilde{\varepsilon}}(l\tau_n/h)-Q_{\widetilde{\varepsilon}}(\tau_n/h)}{Q_{\widetilde{\varepsilon}}(m\tau_n/h)-Q_{\widetilde{\varepsilon}}(\tau_n/h)}\longrightarrow\frac{\log(l)}{\log(m)}.$$

Hence,

$$\frac{\widehat{\gamma}(m\tau_n)-\widehat{\gamma}(\tau_n)}{\gamma(m\tau_n)-\gamma(\tau_n)}\xrightarrow{p}\frac{\log(l)}{\log(m)},$$

which in turn establishes that $\widetilde{G}_n\xrightarrow{p}G/\log(m)$. Combined with Step 1 and (C.10), this shows that

$$\sqrt{\tau_n n}(\widehat{\delta}-\delta)\xrightarrow{d}\mathcal{N}(0,\Omega_\delta),$$

where $\Omega_\delta=(G'WG)^{-1}G'W(I_J\otimes\Delta)\Gamma(L\otimes\Omega_0)\Gamma'(I_J\otimes\Delta')WG(G'WG)^{-1}$. The optimal weighting matrix is then (see, e.g., Wooldridge, 2002, Problem 8.5)

$$W_\delta^*=\left[(I_J\otimes\Delta)\Gamma(L\otimes\Omega_0)\Gamma'(I_J\otimes\Delta')\right]^{-1},$$

and the corresponding asymptotic variance is $\Omega_\delta^*=(G'W_\delta^*G)^{-1}$.

3. Asymptotic normality of $\widehat{\beta}$.

Consider first $\widehat{\beta}_j=-\widehat{\beta}(l_j\tau_n)+\widehat{\gamma}(l_j\tau_n)\widehat{\delta}$ for $j\in\{0,...,J\}$. We have

$$\begin{aligned}\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)}(\widehat{\beta}_j-\beta)=\frac{\gamma(l_j\tau_n)}{\gamma(\tau_n)}\Bigg[&\sqrt{\tau_n n}\left(\frac{\widehat{\gamma}(l_j\tau_n)-\gamma(l_j\tau_n)}{\gamma(l_j\tau_n)}\right)\hat{\delta}+\sqrt{\tau_n n}(\widehat{\delta}-\delta)\\&-\frac{\sqrt{\tau_n n}}{\gamma(l_j\tau_n)}\left(\widehat{\beta}(l_j\tau_n)-\beta(l_j\tau_n)\right)\Bigg].\end{aligned}\tag{C.17}$$

By Lemma D.1 in Appendix D, $\gamma(.)\in RV_0(0)$. Thus, the first ratio on the right-hand side tends to one. We now show that the first and third term in the brackets are $o_P(1)$. We have

$$\sqrt{\tau_n n}\frac{\widehat{\gamma}(l_j\tau_n)-\gamma(l_j\tau_n)}{\gamma(l_j\tau_n)}=\left(\frac{\gamma(m\tau_n)-\gamma(l_j\tau_n)}{\gamma(l_j\tau_n)}\right)\left(e_1'\alpha_n(\widehat{\theta}(l_j\tau_n)-\theta(l_j\tau_n))\right).\tag{C.18}$$

The second term of the right-hand side is $e_1'\widehat{Z}(l_j)$ and is therefore bounded in probability uniformly over j. Because $\gamma(.)\in RV_0(0)$, the first term converges to 0. Thus, the first term in the brackets of the right-hand side of (C.17) is a $o_P(1)$. The same reasoning applies to the third term in the brackets in (C.17).

Hence,

$$\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)}(\widehat{\beta}_j-\beta)=\sqrt{\tau_n n}(\widehat{\delta}-\delta)+o_P(1).$$

Now, because $\widehat{\beta} = \sum_{j=0}^{J} \widehat{\beta}_j/(J+1)$, we obtain

$$\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Omega_\delta).$$

It also follows from the fact that the left-hand side of (C.18) converges to 0 that $\widehat{\gamma(l_j\tau_n)}/\gamma(l_j\tau_n) \xrightarrow{p} 1$, and in particular $\widehat{\gamma(\tau_n)}/\gamma(\tau_n) \xrightarrow{p} 1$. This implies that $\gamma(\tau_n)$ can be replaced by $\widehat{\gamma(\tau_n)}$ in the equation above.

4. Consistency of $\widehat{\beta}$ and polynomial rates of convergence.

First, suppose that $f(u) = o(|u|^{-a})$ as $u \to -\infty$, for some $a > 1$. Fix $\alpha \in (0,1)$ such that $a(1-\alpha) > 1$ and let $\tau_n = \tau_n^*$ be defined by (3.4). As shown in the discussion before Theorem 3.1, such a τ_n satisfies Assumption 9. Moreover, since $\gamma(\tau_n) = G^{-1}(1/n)$ (with $G(\gamma) = F_{\widetilde{\varepsilon}}(\gamma) f(\gamma)^{2(1-\alpha)}$),

$$\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} = \frac{\sqrt{h}}{f^{1-\alpha}(G^{-1}(1/n))G^{-1}(1/n)}.$$

Because $f^{1-\alpha}(u)u \to 0$ as $u \to -\infty$ and $G^{-1}(1/n) \to -\infty$, we get

$$\lim_{n\to\infty} \frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} = -\infty.$$

Thus, $\widehat{\beta}$ is consistent with such a choice of τ_n.

Now, suppose that $f(u) = o(F_{\widetilde{\varepsilon}}(u)^a)$ for some $a > 0$. Consider in this case $\tau_n = n^{-1/(2a+1)}$. Then $\tau_n \to 0$ and $n\tau_n \to \infty$. Because $f(\gamma(\tau)) = o(\tau^a)$, we also have

$$\sqrt{\tau_n n} f(\gamma(\tau_n)) = n^{a/(2a+1)} o\left(n^{-a/(2a+1)}\right) = o(1).$$

Hence, this choice of τ_n satisfies Assumption 9. Besides, $\gamma(.) \in RV_0(0)$. This implies that for any $\alpha > 0$, $|\gamma(\tau_n)| < \tau_n^{-\alpha}$ for n large enough. Choose $0 < \alpha < a$. Then, for n large enough,

$$\frac{\sqrt{\tau_n n}}{|\gamma(\tau_n)|} > n^{(a-\alpha)/(2a+1)}.$$

This ensures that $\widehat{\beta}$ has a polynomial rate of convergence. With such a τ_n, the rate of convergence of $\widehat{\delta}$ is $n^{a/(2a+1)}$, which is also polynomial. This concludes the proof of Theorem 3.1.

C.7 Verification of (2.4) and (3.5) for several copulas

Case 1: Gaussian copula with $\rho > 0$. We just check (3.5), which is stronger than (2.4). We have, after some algebra,

$$\begin{aligned}
1 - \partial_1 C_\rho(u,v) &= 1 - \frac{1}{\varphi(\Phi^{-1}(u))} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(-(\Phi^{-1}(u)^2 - 2\rho\Phi^{-1}(u)s + s^2)/[2(1-\rho^2)])ds \\
&= 1 - \frac{e^{-(1-\rho^2)\Phi^{-1}(u)^2/[2(1-\rho^2)]}}{\sqrt{2\pi}\varphi(\Phi^{-1}(u))} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{\sqrt{(2\pi)1-\rho^2}} \exp\left(-(s-\rho\Phi^{-1}(u))^2/[2(1-\rho^2)]\right) ds \\
&= \Phi\left(\frac{\rho\Phi^{-1}(u) - \Phi^{-1}(v)}{\sqrt{1-\rho^2}}\right).
\end{aligned}$$

Thus, because $\rho > 0$,

$$\sup_{v\in[\underline{v},1], u\leq\tau} 1 - \partial_1 C_\rho(u,v) = \Phi\left(\frac{\rho\Phi^{-1}(\tau) - \Phi^{-1}(\underline{v})}{\sqrt{1-\rho^2}}\right).$$

Now, as $x \to -\infty$, we have $\Phi(x) \sim -\varphi(x)/x$. Because for any $K > 0$, $\exp(-Kx^2) \leq -1/x \leq 1$ for x small enough, we have $\varphi(x/\sigma) \leq \Phi(x) \leq \varphi(x)$ for any $0 < \sigma < 1$. This also implies that $\Phi^{-1}(\tau) \leq \sigma\varphi^{-1}(\tau)$, for τ small enough and with φ^{-1} the inverse of φ on $(-\infty, 0]$. Similarly, for any $m > 0$, there exists $\sigma > 1$ such that for any x small enough, $\varphi(x+m) \leq \varphi(x/\sigma)$. Combining these inequalities, we obtain, for any $K < \rho/\sqrt{1-\rho^2}$,

$$f_C(\tau) \leq \varphi(K\varphi^{-1}(\tau)) = K'\varphi(\Phi^{-1}(\tau))^{K^2} \leq \sqrt{2\pi}^{K^2-1}\tau^{K^2}.$$

The result follows.

Case 2: Archimedean copulas with $\lim_{u\to 0}\Psi(u) = +\infty$ and $\Psi \in RV_\alpha(0)$ with $\alpha \in (0, +\infty]$. Because Ψ is decreasing, we have, by Proposition 0.8 of Resnick (1987), $\Psi^{-1} \in RV_{1/\alpha}(\infty)$. As a result, for all $v \in [\underline{v}, 1]$,

$$u \geq C(u,v) \geq \Psi^{-1}(\Psi(u) + \Psi(\underline{v})) \sim \Psi^{-1}(\Psi(u)) = u \text{ as } u \to 0.$$

In other words,

$$\lim_{u\to 0}\sup_{v\in[\underline{v},1]} |C(u,v)/u - 1| = 0.$$

This implies that

$$\sup_{v\in[\underline{v},1]} \left|\frac{\Psi'(u)}{\Psi'(C(u,v))} - 1\right| = \left|\frac{\Psi'(u)}{\Psi'(l(u)u)} - 1\right|. \tag{C.19}$$

for some function $l(.)$ tending to one as $u \to 0$. Now, by Proposition 0.7 of Resnick (1987),

$\Psi' \in RV_{\alpha-1}(0)$. This implies that the left-hand side of (C.19) tends to 0. (2.4) follows by remarking that $\partial_1 C(u,v) = \Psi'(u)/\Psi' \circ C(u,v)$.

Case 3: Gumbel copulas with $\theta > 1$. Some algebra yields

$$\partial_1 C(u,v) = \frac{1}{1+\Psi(v;\theta)/\Psi(u;\theta)} \frac{C(u,v)\log C(u,v)}{u \log u}.$$

Now, by the fact that $x\log(x)$ is decreasing when x is close to 0 and $C(u,v) \leq u$, we have $C(u,v)\log C(u,v) \geq u\log(u)$, i.e. $\frac{C(u,v)\log C(u,v)}{u\log u} \leq 1$. Because $v \mapsto C(u,v)$ is increasing, $C(u,v)\log C(u,v) \leq C(u,\underline{v})\log C(u,\underline{v})$. Furthermore, $0 \leq \Psi(v,\theta) \leq \Psi(\underline{v},\theta)$. Therefore, we have

$$\begin{aligned}
&\sup_{v\in[\underline{v},1]} |\partial_1 C(u,v) - 1| \\
\leq &\sup_{v\in[\underline{v},1]} \left(\left| \frac{C(u,v)\log C(u,v)}{u\log u} - 1 \right| + \left| \partial_1 C(u,v) - \frac{C(u,v)\log C(u,v)}{u\log u} \right| \right) \\
\leq &\sup_{v\in[\underline{v},1]} \left(1 - \frac{C(u,v)\log C(u,v)}{u\log u} \right) + \sup_{v\in[\underline{v},1]} \left(\frac{\Psi(v,\theta)}{\Psi(v,\theta)+\Psi(u,\theta)} \right) \frac{C(u,v)\log C(u,v)}{u\log u} \\
\leq &(1 - \frac{C(u,\underline{v})\log C(u,\underline{v})}{u\log u}) + \frac{\Psi(\underline{v},\theta)}{\Psi(\underline{v},\theta)+\Psi(u,\theta)}
\end{aligned}$$

$\Psi(u,\theta) \to \infty$ as $u \to 0$, so the second term also converges 0. Therefore, to prove (2.4), it suffices to show that $C(u,\underline{v}) \sim u$. We have, for $\theta > 1$,

$$\begin{aligned}
C(u,\underline{v}) &= \exp\left[-\left((-\log u)^\theta + (-\log \underline{v})^\theta \right)^{1/\theta} \right] \\
&= \exp\left[\log u \left(1 + \left(\frac{-\log \underline{v}}{-\log u} \right)^\theta \right)^{1/\theta} \right] \\
&= \exp\left[\log u + \frac{(-\log \underline{v})^\theta}{\theta(-\log u)^{\theta-1}} + o\left(\frac{1}{(-\log u)^{\theta-1}} \right) \right] \\
&\sim u.
\end{aligned}$$

Case 4: Clayton copula with $\theta > 0$. We obtain in this case

$$1 - \partial_1 C(u,v;\theta) \leq K u^\theta \left(\frac{1}{v^\theta} - 1 \right)$$

Hence, $f_C(\tau) \leq K'\tau^\theta$, where $K' = K\left(\frac{1}{\underline{v}^\theta} - 1 \right)$. (3.5) follows.

Case 5: Rotated Gumbel-Barnett copula with $\theta \in (0,1]$. We have

$$1-\partial_1 C(u,v;\theta) = (1-v)\exp(-\theta\log(u)\log(1-v))(1-\theta\log(1-P)) \leq O\left(u^{-\theta\log(1-\underline{v})}\right) \quad \text{(C.20)}$$

It follows that (3.5) holds.

Case 6: $C(u,v;\theta) = (1+[(u^{-1}-1)^\theta + (v^{-1}-1)^\theta]^{1/\theta})^{-1}$ with $\theta > 1$. In this case,

$$1-\partial_1 C(u,v;\theta) = 1 - \left(\frac{1}{u+[(1-u)^\theta + u^\theta(v^{-1}-1)^\theta]^{1/\theta}}\right)^2 \left[1+\left(\frac{v^{-1}-1}{u^{-1}-1}\right)^\theta\right]^{1/\theta-1} \leq Ku.$$

(3.5) follows.

Case 7: $C(u,v;\theta) = (1+[(u^{-1/\theta}-1)^\theta + (v^{-1/\theta}-1)^\theta]^{1/\theta})^{-\theta}$ with $\theta \geq 1$. We have

$$\begin{aligned}\partial_1 C(u,v;\theta) &= 1 - \left(u^{1/\theta} + [(1-u^{1/\theta})^\theta + u(v^{-1/\theta}-1)^\theta]^{1/\theta}\right)^{-\theta-1} \left[1+\left(\frac{v^{-\frac{1}{\theta}}-1}{u^{-\frac{1}{\theta}}-1}\right)^\theta\right]^{1/\theta-1} \\ &\leq Ku^{1/\theta}\end{aligned}$$

which implies (3.5).

Case 8: $C(u,v;\theta) = \theta/\log(\exp(\theta/u)+\exp(\theta/v)-\exp(\theta))$ with $\theta > 0$. We have

$$\begin{aligned}1-\partial_1 C(u,v;\theta) &= 1 - 1/(1+\log(1+(\exp(\theta/v)-\exp(\theta))\exp(-\theta/u)))^2 \frac{1}{1+(\exp(\theta/v)-\exp(\theta))\exp(-\theta/u)} \\ &\leq K\exp(-\theta/u)\end{aligned}$$

Thus Condition (3.5) is easily satisfied. In this case, any polynomial rate slower than the parametric rate is in fact possible.

Case 9: $C(u,v;\theta) = [\log(\exp(u^{-\theta})+\exp(v^{-\theta})-e)]^{-1/\theta}$ with $\theta > 0$. Start from

$$\begin{aligned}1-\partial_1 C(u,v;\theta) &= 1 - \left[1+u^\theta\log\left(1+\frac{\exp(v^{-\theta})-e}{\exp(u^{-\theta})}\right)\right]^{-1/\theta-1} \frac{1}{1+\frac{\exp(v^{-\theta})-e}{\exp(u^{-\theta})}} \\ &\leq K_1 u^\theta \log\left(1+[\exp(v^{-\theta})-e]\exp(-u^{-\theta})\right) + K_2\exp(-u^{-\theta}) \\ &\leq K\exp(-u^{-\theta})\end{aligned}$$

Therefore, Condition (3.5) is easily satisfied and once more, any polynomial rate slower than parametric rate is possible.

C.8 Proof of Theorem 3.2

We use the notations of the proof of Theorem 3.1 (see Section C.6) along with the notations introduced in the text before Theorem 3.2. First consider δ^1. Let

$$G_{2n} = -\frac{\partial g_{2n}(\delta^1)}{\partial \delta^1} = \begin{pmatrix} \widehat{\gamma}(l_1\tau_n) - \widehat{\gamma}(\tau_n) \\ \vdots \\ \widehat{\gamma}(l_J\tau_n) - \widehat{\gamma}(\tau_n) \end{pmatrix} \otimes I_{d_\delta}.$$

Reasoning as in the proof of Theorem 3.1, above Equation (C.10),

$$\sqrt{\tau_n n}(\widehat{\delta}^1 - \delta^1) = \left[G_{2n}' W_{2n} G_{2n}\right]^{-1} G_{2n}' W_{2n} (I_J \otimes \widetilde{\Psi}\Delta)\Gamma \widehat{Z}_n(l_1, ..., l_J).$$

Besides, one can show that $G_{2n} \xrightarrow{p} G_\delta / \log(m)$ as in the proof of Theorem 3.1. Combined with the asymptotic normality of $\widehat{Z}_n(\tau)$, this implies that

$$\sqrt{\tau_n n}(\widehat{\delta}^1 - \delta^1) \xrightarrow{d} \mathcal{N}\left(0, \left[G_\delta' W_2 G_\delta\right]^{-1} G_\delta' W_2 (I_J \otimes \widetilde{\Psi}\Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'\widetilde{\Psi}')W_2 G_\delta \left[G_\delta' W_2 G_\delta\right]^{-1}\right).$$

The optimal weighting matrix is then

$$W_{\delta^1}^* = \left[(I_J \otimes \widetilde{\Psi}\Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'\widetilde{\Psi}')\right]^{-1}$$

and the corresponding asymptotic variance is $(G_\delta' W_{\delta^1}^* G_\delta)^{-1}$.

Next, we derive the asymptotic properties of $\widehat{\beta}^1$. Reasoning as previously, we have

$$\log(m)\alpha_n(\widehat{\beta}^1 - \beta^1) = (G_\beta' W_1 G_\beta)^{-1} G_\beta' W_1 (\Gamma_3 \otimes \Psi\Gamma_2) \log(m) \widehat{Z}_n(l_1, ..., l_J) + o_P(1),$$

Remark that $\log(m)\alpha_n = \sqrt{n\tau_n}\lambda_n / \gamma(\tau_n)$, with $\lambda_n = \gamma(\tau_n)\log(m)/[\gamma(m\tau_n) - \gamma(\tau_n)]$. Then the asymptotic normality of $\widehat{Z}_n(l_1, ..., l_J)$ yields

$$\lambda_n \frac{\sqrt{\tau_n n}}{\gamma(\tau_n)}(\widehat{\beta}^1 - \beta^1) \xrightarrow{d} \mathcal{N}\left(0, (G_\beta' W_1 G_\beta)^{-1} G_\beta' W_1 (\Gamma_3 \otimes \Psi\Gamma_2)(L \otimes \Omega_0)(\Gamma_3' \otimes \Gamma_2'\Psi')W_1 G_\beta (G_\beta' W_1 G_\beta)^{-1}\right).$$

It follows from the proof of Theorem 3.1 that $\widehat{\lambda}_n / \lambda_n \xrightarrow{p} 1$ and $\widehat{\gamma(\tau_n)}/\gamma(\tau_n) \xrightarrow{p} 1$. Therefore, we can replace λ_n by $\widehat{\lambda}_n$ and $\gamma(\tau_n)$ by $\widehat{\gamma(\tau_n)}$ in the previous equation. Finally, the optimal matrix is

$$W_{\beta^1}^* = \left[(\Gamma_3 \otimes \Psi\Gamma_2)(L \otimes \Omega_0)(\Gamma_3' \otimes \Gamma_2'\Psi')\right]^{-1}$$

and the corresponding asymptotic variance is $(G_\beta' W_{\beta^1}^* G_\beta)^{-1}$.

D Technical lemmas

Lemma D.1. *If Assumption 3 (ii)-(iii) hold, then S_ε is rapidly varying at $+\infty$, i.e. its extreme value index is 0. Moreover, $Q_{\widetilde{\varepsilon}} \in RV_0(0)$.*

Proof. Because $\sup(\text{Supp}(\varepsilon)) = \infty$, S_ε is not in the attraction domain of type III extreme value distributions (see Resnick, 1987, Proposition 1.13). Suppose S_ε is not rapidly varying. Then, S_ε is not either in the attraction domain of type I extreme value distribution (See Resnick, 1987, Exercise 1.1.9). So S_ε is in the attraction domain of type II extreme value distribution, i.e. $S_\varepsilon \in RV_{-\xi^{-1}}(+\infty)$ with extreme value index $\xi > 0$. We also have

$$\frac{S_{\exp(\varepsilon)}(tx)}{S_{\exp(\varepsilon)}(x)} = \frac{S_\varepsilon(u(x)\log(x))}{S_\varepsilon(\log(x))} \tag{D.1}$$

where $u(x) = \frac{\log(t)+\log(x)}{\log(x)} \to 1$ as $x \to +\infty$. Because $S_\varepsilon \in RV_{-\xi^{-1}}(+\infty)$, the right-hand side of Equation (D.1) converges to 1. This implies that $S_{\exp(\varepsilon)}$ is slowly varying, a contradiction. Thus, S_ε is rapidly varying at $+\infty$.

To prove the second result, note that $1/S_\varepsilon$ is nondecreasing, rapidly varying at $+\infty$ and satisfies $1/S_\varepsilon(+\infty) = +\infty$. Thus, by Proposition 0.8 of Resnick (1987), $(1/S_\varepsilon)^{\leftarrow} \in RV_0(\infty)$. Remark that $(1/S_\varepsilon)^{\leftarrow}(1/\tau) = -Q_{\widetilde{\varepsilon}}(\tau)$. Hence, $Q_{\widetilde{\varepsilon}} \in RV_0(0)$. □

Lemma D.2. *Suppose that Assumptions 3 (ii)- (iii) and 8 hold. Then $Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau) \in RV_0(0)$, $Q'_{\widetilde{\varepsilon}} \in RV_{-1}(0)$ and for any positive (l, m),*

$$\lim_{\tau\to 0} \frac{Q_{\widetilde{\varepsilon}}(l\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} = \frac{\log(l)}{\log(m)}.$$

Proof. We first prove the last point. By Lemma D.1, F_ε is in the attraction domain of type I distribution. Then by Proposition 0.10 in Resnick (1987), $\tau \mapsto -Q_{\widetilde{\varepsilon}}(\tau)$ is Π-varying with auxiliary function $\tau \mapsto Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)$, namely

$$\lim_{\tau\to 0} \frac{Q_{\widetilde{\varepsilon}}(l\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} = \log(l) \tag{D.2}$$

for all $l > 0$. Then

$$\frac{Q_{\widetilde{\varepsilon}}(l\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} = \frac{Q_{\widetilde{\varepsilon}}(l\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \Big/ \frac{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \to \frac{\log(l)}{\log(m)}.$$

Turning to the first point, we have, by (D.2),

$$\frac{Q_{\widetilde{\varepsilon}}(ex\tau) - Q_{\widetilde{\varepsilon}}(x\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} = \frac{Q_{\widetilde{\varepsilon}}(ex\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} - \frac{Q_{\widetilde{\varepsilon}}(x\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \to \log(ex) - \log(x) = 1.$$

Finally, let us prove the second point. By monotonicity of $Q'_{\widetilde{\varepsilon}}$,

$$\frac{Q'_{\widetilde{\varepsilon}}(b\tau)\tau(b-a)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \geq \frac{Q_{\widetilde{\varepsilon}}(b\tau) - Q_{\widetilde{\varepsilon}}(a\tau)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \geq \frac{Q'_{\widetilde{\varepsilon}}(a\tau)\tau(b-a)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)},$$

for any $b > a > 0$. Therefore, using (D.2),

$$\limsup_{\tau\to 0} \frac{Q'_{\widetilde{\varepsilon}}(a\tau)\tau}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \leq \frac{\log(b) - \log(a)}{b-a}.$$

Letting $b \downarrow a$, we obtain

$$\limsup_{\tau\to 0} \frac{Q'_{\widetilde{\varepsilon}}(a\tau)\tau}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \leq \frac{1}{a},$$

for any $a > 0$. Similarly, we obtain from the other inequality

$$\liminf_{\tau\to 0} \frac{Q'_{\widetilde{\varepsilon}}(b\tau)\tau}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \geq \frac{1}{b},$$

for any $b > 0$. By letting $a = b = 1$, we obtain

$$Q'_{\widetilde{\varepsilon}}(\tau) \sim \frac{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)}{\tau} \tag{D.3}$$

This, combined with $Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau) \in RV_0(0)$, shows the second point. □

Lemma D.3. *Suppose that Assumptions 1-4 and 8 hold. Then, for all $x \in Supp(X)$,*

$$\lim_{\tau\to 0} \left| \frac{Q_{\breve{\varepsilon}|X}(\tau|x) - Q_{\widetilde{\varepsilon}}(\tau/h)(1 + x'\delta)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \right| = 0,$$

$$f_{\breve{\varepsilon}|X}\left(Q_{\breve{\varepsilon}|X}(\tau|x)\right) \sim h f_{\widetilde{\varepsilon}}(Q_{\widetilde{\varepsilon}}(\tau/h))/(1 + x'\delta). \tag{D.4}$$

Proof. For the first point, fix $\Delta \in (0, h)$ and remark first that by Lemma D.2,

$$\begin{aligned} \lim_{\tau\to 0} \frac{Q_{\widetilde{\varepsilon}}(\tau/(h+\Delta)) - Q_{\widetilde{\varepsilon}}(\tau/h)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} &= \lim_{\tau\to 0} \left[\frac{Q_{\widetilde{\varepsilon}}(\tau/(h+\Delta)) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} - \frac{Q_{\widetilde{\varepsilon}}(\tau/h) - Q_{\widetilde{\varepsilon}}(\tau)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \right] \\ &\to \frac{-\log(h+\Delta)}{\log(m)} + \frac{\log(h)}{\log(m)} = \frac{\log[h/(h+\Delta)]}{\log(m)} \end{aligned} \tag{D.5}$$

and the same holds replacing Δ by $-\Delta$.

Besides, by definition of the quantiles of $\breve{\varepsilon}|X=x$, we have, for all τ small enough,

$$\begin{aligned}\tau &\leq P(\breve{\varepsilon} \leq Q_{\breve{\varepsilon}|X}(\tau|x)|X=x)\\ &= P(Y \geq x'\beta - Q_{\breve{\varepsilon}|X}(\tau|x)|X=x)\\ &= P(Y^* \geq x'\beta - Q_{\breve{\varepsilon}|X}(\tau|x), D=1|X=x)\\ &= \int_{x'\beta - Q_{\breve{\varepsilon}|X}(\tau|x)}^{\infty} P(D=1|Y^*=y, X=x) dP^{Y^*|X=x}(y).\end{aligned}$$

For τ small enough, $P(D=1|Y^*=y, X=x) \in [h-\Delta, h+\Delta]$ for all $y > x'\beta - Q_{\breve{\varepsilon}|X}(\tau|x)$. Thus,

$$\tau \leq (h+\Delta) P\left[\widetilde{\varepsilon}(1+x'\delta) \geq Q_{\breve{\varepsilon}|X}(\tau|x)\right].$$

Similarly, using $\tau \geq (h-\Delta)P(\breve{\varepsilon} < Q_{\breve{\varepsilon}|X}(\tau|x)|X=x)$,

$$\tau \geq (h-\Delta) P\left[\widetilde{\varepsilon}(1+x'\delta) \geq Q_{\breve{\varepsilon}|X}(\tau|x)\right].$$

Then, by definition of the quantiles of $\widetilde{\varepsilon}$,

$$(1+x'\delta)Q_{\widetilde{\varepsilon}}(\tau/(h+\Delta)) \leq Q_{\breve{\varepsilon}|X}(\tau|x) \leq (1+x'\delta)Q_{\widetilde{\varepsilon}}(\tau/(h-\Delta)).$$

This, together with Equation (D.5)

$$\begin{aligned}&\limsup_{\tau} \left| \frac{Q_{\breve{\varepsilon}|X}(\tau|x) - Q_{\widetilde{\varepsilon}}(\tau/h)(1+x'\delta)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)} \right|\\ \leq &(1+x'\delta) \limsup_{\tau} \frac{\max\left(Q_{\widetilde{\varepsilon}}(\tau/(h-\Delta)) - Q_{\widetilde{\varepsilon}}(\tau/h),\, Q_{\widetilde{\varepsilon}}(\tau/h) - Q_{\widetilde{\varepsilon}}(\tau/(h+\Delta))\right)}{Q_{\widetilde{\varepsilon}}(m\tau) - Q_{\widetilde{\varepsilon}}(\tau)}\\ \leq &(1+x'\delta) \frac{\max\left(\log(h/(h-\Delta)), \log((h+\Delta)/h)\right)}{\log(m)}.\end{aligned}$$

By letting Δ tend to 0, the left-hand side tends to zero. The first result follows.

Now let us turn to the second result. We first show that for any fixed x, $Q_{\breve{\varepsilon}|X}(\tau|x)$ is $\Pi-$varying. We have

$$\begin{aligned}\frac{Q_{\breve{\varepsilon}|X}(m\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} =& \frac{Q_{\breve{\varepsilon}|X}(m\tau|x) - (1+x'\delta)Q_{\widetilde{\varepsilon}}(m\tau/h)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)} - \frac{Q_{\breve{\varepsilon}|X}(\tau|x) - (1+x'\delta)Q_{\widetilde{\varepsilon}}(\tau/h)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)}\\ &+ (1+x'\delta)\frac{Q_{\widetilde{\varepsilon}}(m\tau/h) - Q_{\widetilde{\varepsilon}}(\tau/h)}{Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)}\end{aligned}$$

By Lemma D.2, $\tau \mapsto Q_{\widetilde{\varepsilon}}(e\tau) - Q_{\widetilde{\varepsilon}}(\tau)$ is slowly varying. Thus, by the first result of this lemma, the first and second term converge to zero. Since $Q_{\widetilde{\varepsilon}}(\tau)$ is $\Pi-$varying, the third term

converges to $(1 + x'\delta)\log(m)$. Therefore

$$\frac{Q_{\breve{\varepsilon}|X}(m\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)}{Q_{\tilde{\varepsilon}}(e\tau) - Q_{\tilde{\varepsilon}}(\tau)} \sim (1 + x'\delta)\log(m).$$

Then

$$\frac{Q_{\breve{\varepsilon}|X}(m\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)}{Q_{\breve{\varepsilon}|X}(e\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)} = \frac{Q_{\breve{\varepsilon}|X}(m\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)}{Q_{\tilde{\varepsilon}}(e\tau) - Q_{\tilde{\varepsilon}}(\tau)} \frac{Q_{\tilde{\varepsilon}}(e\tau) - Q_{\tilde{\varepsilon}}(\tau)}{Q_{\breve{\varepsilon}|X}(e\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)} \to \log(m),$$

which proves that $Q_{\breve{\varepsilon}|X}(.|x)$ is $\Pi-$varying. Now, remark that for y small enough,

$$\begin{aligned} P(\breve{\varepsilon} \leq y|X = x) &= P(\tilde{Y} + X'\beta \leq y|X = x) \\ &= P(Y^* \geq -y + x'\beta, D = 1|X = x) \\ &= P\left(\tilde{\varepsilon} \leq \frac{y - x'\beta}{1 + x'\delta}|D = 1, X = x\right) P(D = 1|X = x). \end{aligned}$$

This equality, combined with Assumption 8 and the fact that X is bounded, ensures that the cdf of $\breve{\varepsilon}|X$ is increasing. As a result, $Q'_{\breve{\varepsilon}|X}(.|x)$ is decreasing at the lower tail and we have, by the same reasoning as in Lemma D.2,

$$Q_{\breve{\varepsilon}|X}(\tau|x)' \sim \tau(Q_{\breve{\varepsilon}|X}(m\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x)). \tag{D.6}$$

Combining Equations (D.3) and (D.6), we obtain

$$\frac{Q_{\tilde{\varepsilon}}(\tau/h)'}{Q_{\breve{\varepsilon}|X}(\tau|x)'} \sim \frac{(Q_{\tilde{\varepsilon}}(m\tau) - Q_{\tilde{\varepsilon}}(\tau))}{h(Q_{\breve{\varepsilon}|X}(m\tau|x) - Q_{\breve{\varepsilon}|X}(\tau|x))} \sim \frac{1}{h(1 + x'\delta)}$$

This proves the second result of the lemma. □

Lemma D.4. *Suppose that Assumptions 1 - 9 hold and let $\Lambda_n(z, \tau)$ be defined as in* (C.12)*. Then*

$$\Lambda_n(z, \tau_n) \xrightarrow{p} \frac{1}{2}\log(m) z' \mathcal{Q}_X z.$$

Proof. By Lemma 9.6 in Chernozhukov (2005), the variance of $\Lambda_n(z, \tau)$ converges to 0. Thus it suffices to prove that $E[\Lambda_n(z, \tau_n)] \to \frac{1}{2}\log(m) z' \mathcal{Q}_H z$. Let us define, for any $(s, t) \in \mathbb{R}^2$,

$$m(s,t) = \begin{cases} 1 & \text{if } 0 < s \leq t, \\ -1 & \text{if } t \leq s < 0, \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$
\begin{aligned}
&E\left[\Lambda_n(z,\tau_n)\right] \\
=&\frac{\alpha_n}{\sqrt{\tau_n n}} nE\left[\int_0^{(z_1+X'z_2)/\alpha_n} \mathbb{1}\{\tilde{Y}-\gamma(\tau_n)-X'\beta(\tau_n)\le s\}-\mathbb{1}\{\tilde{Y}-\gamma(\tau_n)-X'\beta(\tau_n)\le 0\}ds\right] \\
=&\frac{n}{\sqrt{\tau_n n}} E\left[\int_0^{z_1+X'z_2} \mathbb{1}\{\tilde{Y}-\gamma(\tau_n)-X'\beta(\tau_n)\le s/\alpha_n\}-\mathbb{1}\{\tilde{Y}-\gamma(\tau_n)-X'\beta(\tau_n)\le 0\}ds\right] \\
=&\frac{n}{\sqrt{\tau_n n}} E\left[\int_0^{z_1+X'z_2} \mathbb{1}\{\breve{\varepsilon}-(1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h)\le s/\alpha_n\}-\mathbb{1}\{\breve{\varepsilon}-(1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h)\le 0\}ds\right] \\
=&nE\left[\int_0^{z_1+X'z_2} \frac{F_{\breve{\varepsilon}|X}((1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h)+s/\alpha_n)-F_{\breve{\varepsilon}|X}((1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h))}{\sqrt{\tau_n n}}ds\right] \\
=&E\left[\int_{-\infty}^{+\infty} m(s,z_1+X'z_2)s\,\frac{nf_{\breve{\varepsilon}|X}\left[(1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h)+V_s\right]}{\alpha_n\sqrt{\tau_n n}}ds\right], \qquad \text{(D.7)}
\end{aligned}
$$

where for each s, V_s is a random variable satisfying $V_s \in [0, s/\alpha_n]$. Let

$$
U_n(s) = m(s,z_1+X'z_2)s\,\frac{nf_{\breve{\varepsilon}|X}\left[(1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h)+V_s\right]}{\alpha_n\sqrt{\tau_n n}}.
$$

We first show that

$$
U_n(s) \xrightarrow{p.s.} \frac{m(s,z_1+X'z_2)s\,\log(m)}{1+X'\delta}. \qquad \text{(D.8)}
$$

Since $1/\alpha_n = o(Q_{\widetilde{\varepsilon}}(m\tau_n)-Q_{\widetilde{\varepsilon}}(\tau_n))$, we have $V_s = o(Q_{\widetilde{\varepsilon}}(m\tau_n)-Q_{\widetilde{\varepsilon}}(\tau_n)))$. Moreover, by Lemma D.3,

$$
Q_{\breve{\varepsilon}|X}(\tau_n|x) - Q_{\widetilde{\varepsilon}}(\tau_n/h)(1+x'\delta) = o(Q_{\widetilde{\varepsilon}}(m\tau_n)-Q_{\widetilde{\varepsilon}}(\tau_n)).
$$

Then, following the same argument as Chernozhukov (2005) after his Equation (9.57),

$$
\frac{f_{\breve{\varepsilon}|X}\left[(1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h)+V_s\right]}{f_{\breve{\varepsilon}|X}\left(Q_{\breve{\varepsilon}|X}(\tau_n|x)\right)} \xrightarrow{p} 1. \qquad \text{(D.9)}
$$

Besides, by Lemma D.3,

$$
f_{\breve{\varepsilon}|X}\left(Q_{\breve{\varepsilon}|X}(\tau_n|x)\right) \sim hf_{\widetilde{\varepsilon}}(Q_{\widetilde{\varepsilon}}(\tau_n/h))/(1+x'\delta). \qquad \text{(D.10)}
$$

Now, by definition of α_n and because $Q'_{\widetilde{\varepsilon}} \in RV_{-1}(0)$ by Lemma D.2,

$$\begin{aligned}\frac{nhf_{\widetilde{\varepsilon}}(Q_{\widetilde{\varepsilon}}(\tau_n/h))}{\alpha_n\sqrt{\tau_n n}} &= \frac{h(Q_{\widetilde{\varepsilon}}(m\tau_n/h) - Q_{\widetilde{\varepsilon}}(\tau_n/h))f_{\widetilde{\varepsilon}}(Q_{\widetilde{\varepsilon}}(\tau_n/h))}{\tau_n} \\ &= \left[\int_1^m \frac{Q'_{\widetilde{\varepsilon}}(s\tau_n/h)}{Q'_{\widetilde{\varepsilon}}(\tau_n/h)}ds\right] \to \left[\int_1^m \frac{ds}{s}\right] = \log(m), \end{aligned} \tag{D.11}$$

where the second last convergence is because, by Proposition 0.5 of Resnick (1987), $\frac{Q'_{\widetilde{\varepsilon}}(s\tau_n/h)}{Q'_{\widetilde{\varepsilon}}(\tau_n/h)} \to \frac{1}{s}$ locally uniformly. $\frac{1}{s}$ is bounded over $[1, m]$, so dominated convergence theorem can be applied. Combining (D.9), (D.10) and (D.11) proves that (D.8) holds.

Next, we prove that for n large enough,

$$|U_n(s)| \leq U(s), \quad \text{with} \quad E\left(\int_{-\infty}^{\infty} U(s)ds\right) < \infty. \tag{D.12}$$

Together with (D.8), this will allow us to use the dominated convergence theorem on the right-hand side of (D.7). We bound $|U_n(s)|$ for $|s| \leq |z_1 + X'z_2|$, since $m(s, z_1 + X'z_2) = 0$ otherwise.

First, because X is bounded, $\sup_{x\in \text{Supp}(X)} \gamma(\tau_n) + x'\beta(\tau_n) \to -\infty$. Thus, for any $|s| \leq |z_1 + X'z_2|$, we have, for n large enough, $\gamma(\tau_n) + X'\beta(\tau_n) < 0$ and $\gamma(\tau_n) + X'\beta(\tau_n) + s/\alpha_n < 0$. Hence, by definition of $\widetilde{Y}$ and Y^*,

$$\{\widetilde{Y} \in (\gamma(\tau_n)+X'\beta(\tau_n), \gamma(\tau_n)+X'\beta(\tau_n)+s/\alpha_n]\} \subset \{-Y^* \in (\gamma(\tau_n)+X'\beta(\tau_n), \gamma(\tau_n)+X'\beta(\tau_n)+s/\alpha_n]\}.$$

Taking conditional expectations, this implies that for any $|s| \leq |z_1 + X'z_2|$ and n large enough,

$$\frac{|s|}{\alpha_n} f_{\widetilde{\varepsilon}|X}\left[(1 + X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h) + V_s\right] \leq \left|F_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}(\tau_n/h) + \frac{s}{\alpha_n(1 + X'\delta)}\right) - F_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}(\tau_n/h)\right)\right|.$$

By the mean value theorem,

$$\left|F_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}(\tau_n/h) + \frac{s}{\alpha_n(1 + X'\delta)}\right) - F_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}(\tau_n/h)\right)\right| = |s|\frac{f_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}(\tau_n/h) + V'_s\right)}{\alpha_n(1 + X'\delta)}, \tag{D.13}$$

where $V'_s \in [0, s/(\alpha_n(1 + X'\delta))]$. Because $s/(1 + x'\delta)$ is bounded for all $|s| \leq |z_1 + x'z_2|$ and all $x \in \text{Supp}(X)$, $|V'_s| \leq K/\alpha_n$. Now, by Lemma D.2, we have, for any $\eta > 0$,

$$\begin{aligned}\alpha_n\left[Q_{\widetilde{\varepsilon}}((1 + \eta)\tau_n/h) - Q_{\widetilde{\varepsilon}}(\tau_n/h)\right] &= \sqrt{\tau_n n}\frac{Q_{\widetilde{\varepsilon}}((1 + \eta)\tau_n/h) - Q_{\widetilde{\varepsilon}}(\tau_n/h)}{Q_{\widetilde{\varepsilon}}(m\tau_n/h) - Q_{\widetilde{\varepsilon}}(\tau_n/h)} \\ &\sim \sqrt{\tau_n n}\frac{\log(1 + \eta)}{\log(m)} \to \infty.\end{aligned}$$

Hence, for n large enough,

$$Q_{\widetilde{\varepsilon}}((1+\eta)\tau_n/h) \geq Q_{\widetilde{\varepsilon}}(\tau_n/h) + \frac{K}{\alpha_n} \geq Q_{\widetilde{\varepsilon}}(\tau_n/h) + V_s'.$$

Plugging this inequality in (D.13) and using monotonicity of $f_{\widetilde{\varepsilon}}$, we obtain

$$\frac{|s|}{\alpha_n} f_{\check{\varepsilon}|X}\left[(1+X'\delta)Q_{\widetilde{\varepsilon}}(\tau_n/h) + V_s\right] \leq f_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}((1+\eta)\tau_n/h)\right).$$

Because $1 + X'\delta$ is bounded from below, we finally get

$$U_n(s) \leq K|s|\mathbb{1}\{|s| \leq |z_1 + X'z_2|\}\frac{n f_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}((1+\eta)\tau_n/h)\right)}{\alpha_n\sqrt{\tau_n n}}.$$

We have shown in (D.11) that the sequence $n f_{\widetilde{\varepsilon}}\left(Q_{\widetilde{\varepsilon}}(\tau_n/h)\right)/(\alpha_n\sqrt{\tau_n n})$ admits a finite limit. It is therefore bounded. Hence, we finally have, for n large enough, $|U_n(s)| \leq U(s)$ with $U(s) = K|s|\mathbb{1}\{|s| \leq |z_1 + X'z_2|\}$. Thus, (D.12) holds and by the dominated convergence theorem applied to (D.7).

$$E\left[G_n(z,\tau_n)\right] \to E\left[\frac{\log(m)}{1+X'\delta}\int_0^{(z_1+X'z_2)} sds\right] = \frac{1}{2}\log(m) z'\mathcal{Q}_H z.$$

□