3-2018

Extremal quantile regressions for selection models and the black-white wage gap

Xavier D'HAULTFOEUILLE
CREST

Arnaud MAUREL
Duke University

Yichong ZHANG
Singapore Management University, yczhang@smu.edu.sg

Citation

D'HAULTFOEUILLE, Xavier; MAUREL, Arnaud; and ZHANG, Yichong. Extremal quantile regressions for selection models and the black-white wage gap. (2018). *Journal of Econometrics*. 203, (1), 129-142. Research Collection School Of Economics.
Available at: http://ink.library.smu.edu.sg/soe_research/2030

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

Xavier D'Haultfœuille† Arnaud Maurel‡ Yichong Zhang§

July 2017

Abstract

We consider the estimation of a semiparametric sample selection model without instrument or large support regressor. Identification relies on the independence between the covariates and selection, for arbitrarily large values of the outcome. We propose a simple estimator based on extremal quantile regression and establish its asymptotic normality 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 play a key role in explaining the black-white wage gap.

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

JEL codes: C21, C24, J31.

---

*We are grateful to Derek Neal for useful suggestions and providing us with the sample used in his 1996 *JPE* article. We also thank the editor, two referees, Peter Arcidiacono, Victor Chernozhukov, Ivan Fernandez-Val, Shakeeb Khan, Pat Kline, Adam Rosen, Lowell Taylor, Ed Vytlacil, seminar participants at Boston College, CREST, Duke, Institute for Advanced Studies, Paris 1, Toulouse, Universidad Pablo de Olavide (Seville), UT-Austin, and attendees of the 2013 North American Summer Meeting of the Econometric Society (Los Angeles), the 2013 Asian Meeting of the Econometric Society (Singapore), the IAAE 2014 Annual Conference (London), the 2014 Triangle Econometrics Conference and the 2015 World Congress of the Econometric Society (Montreal) for useful comments and suggestions.

†CREST. Address: 5 avenue Henry le Chatelier, 91120 Palaiseau, France; E-mail address: xavier.dhaultfoeuille@ensae.fr.

‡Corresponding author. Duke University, NBER and IZA. Address: Department of Economics, 213 Social Sciences, Durham, NC 27708-0097; Telephone number: (919)660-1851; Fax: (919)681-7984; E-mail address: apm16@duke.edu.

§Singapore Management University. Address: 90 Stamford Rd, Singapore, 178903. E-mail address: yczhang@smu.edu.sg.

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 others, Heckman, 1974, 1979, 1990, Ahn and Powell, 1993, Donald, 1995, Buchinsky, 1998, Chen and Khan, 2003, Das et al., 2003, Newey, 2009 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. In many applications, however, such a regressor is hard to come by.

Starting from this observation, D'Haultfoeuille and Maurel (2013) show that identification in the absence of an instrument or a large support covariate is in fact possible. Their key condition is that selection becomes independent of the covariates at infinity, i.e., when the *outcome* takes arbitrarily large values. The idea behind is that if selection is indeed endogenous, one can expect the effect of the outcome on selection to dominate those of the covariates, for sufficiently large values of the outcome.

This paper builds on this insight and develops a novel inference method for a class of semi-parametric models subject to endogenous selection. Specifically, denoting by Y^* and X_1 the outcome and covariates of interest, and by X_{-1} other covariates (so that the covariates vector is given by $X = (X_1', X_{-1}')'$), we consider the following outcome equation:

$$Y^* = X_1'\beta_1 + \varepsilon$$

where, for any $\tau \in (0, 1)$, the τ-th conditional quantile of ε satisfies $Q_{\varepsilon|X}(\tau|X) = \beta_0(\tau) + X_{-1}'\beta_{-1}(\tau)$. Denoting by D the selection dummy, the econometrician only observes $(D, Y = DY^*, X)$. In this framework, the effect of interest β_1 is identified from the analysis of D'Haultfoeuille and Maurel (2013). In this paper, we extend their result by directly relating β_1 to the upper conditional quantiles of Y. Following this new constructive identification result, we then develop a consistent and asymptotically normal estimator of β_1. We propose an estimator based on extremal quantile regression, that is quantile regression applied to the upper tail of Y (see Chernozhukov et al., 2016, for a survey).[1] Throughout the paper we focus on the intermediate order case, which corresponds to situations where the quantile index goes to one as the sample size tends to infinity, but at a slower rate than the sample size.

Unlike prior estimation methods for sample selection models, we propose a distribution-free estimator that does not require an instrument for selection nor a large support regressor. Besides and importantly, we do not restrict the selection process, apart from the independence at infinity condition mentioned above. In the context of standard selection models, this condition translates into a restriction on the dependence between the error terms of the outcome and selection equation, which is mild provided that selection is indeed endogenous. The structure of the outcome equation, which generalizes the standard location shift model by allowing for heterogeneous effects of the covariates X_{-1} on different parts of the distribution, also plays an important role for identification.[2] Importantly, these assumptions are testable, since they imply that for large quantile indices, the estimators of β_1 obtained using different quantile indices are close.

The main difficulty in establishing the asymptotic properties of our estimator is that, because of selection, extremal conditional quantiles are not exactly linear, but only equivalent to a linear form as the quantile index tends to one. Hence, we face a bias-variance trade-off that is typical in non- or semiparametric analysis. Choosing a moderately large quantile index decreases the variance of the estimator, but this comes at the price of a higher bias. Conversely, choosing a very large 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 for extremal quantile regressions 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.[3] We solve this issue by proving consistency of bootstrap in this context, which is a result of independent interest. While Falk (1991) showed that the bootstrap is valid for unconditional intermediate order quantiles and Chernozhukov and

[1] A Stata code and its user guide are available on the following webpage: `http://www.amaurel.net/Research`.

[2] The location shift specification is very common in the econometrics literature. See, e.g., in the context of sample selection models, Ahn and Powell (1993), Buchinsky (1998) and Newey (2009). Examples of papers using more general location-scale specifications include Chen and Khan (2003) and Chen et al. (2005) in the context of sample selection and censored regression models, respectively.

[3] This is broadly similar to the convergence rates discussed in Andrews and Schafgans (1998), Schafgans and Zinde-Walsh (2002), and Khan and Tamer (2010), the main difference being that the tail behavior of the outcome is going to play a key role here, rather than that of the covariates.

Fernandez-Val (2011) proved the validity of a modified subsampling method for conditional extremal order quantiles, our paper is, to the best of our knowledge, the first to prove the consistency of bootstrap in the context of conditional intermediate order quantiles.

Asymptotic normality and unbiasedness of our estimator require an appropriate choice of the quantile index, similarly to nonparametric kernel regressions that require an appropriate bandwidth choice. We propose a heuristic data-driven procedure that selects the quantile index by minimizing a criterion function capturing the trade-off between bias and variance. In particular, we use subsampling combined with a simple minimum distance estimator to proxy the bias term, which, in this setting, cannot be simply estimated. Monte Carlo simulation results show that our estimator performs well in finite samples. We further provide evidence that bootstrap yields good coverage in small samples, and that our specification test also exhibits good finite sample properties.

We then 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 and 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 and Heckman (1977), correcting for selection is a priori important in order to consistently estimate the black-white differential in terms of potential wages.

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. Our key identifying assumption of independence at infinity, which generalizes the condition imposed in previous work on this question by Neal and Johnson (1996) and Johnson et al. (2000), is natural in this context. For the NLSY79 cohort, we find a smaller residual wage gap (10.6 pp) than the one obtained using the imputation method of Neal and Johnson (1996) and Johnson et al. (2000), which is consistent with our approach being based on a weaker identifying restriction on the selection process.[4] Overall, our estimates confirm the key takeaway of Neal and Johnson (1996) by providing evidence of a major role played by the black-white AFQT gap.

Turning to the evolution across the NLSY79 and NLSY97 cohorts, we find that the black-white wage gap is essentially stable between 1990 and 2007, whether we control or not for premarket factors such as AFQT and family background characteristics. These results suggest

[4]Other noteworthy papers analyzing the black-white wage gap while using imputation methods to correct for selection into the workforce include Brown (1984), Smith and Welch (1989), Juhn (2003), Neal (2004), Neal (2006), and Neal and Rick (2014).

that the lack of progress in closing the black-white wage gap in the last decades, which has recently been documented in the literature (see, e.g, Neal and Rick, 2014), still holds after controlling for differential selection into the workforce.

The remainder of the paper is organized as follows. Section 2 presents the set-up, discusses the identification, defines the estimators and then establishes the main asymptotic normality results. Section 3 discusses the practical implementation of our estimator, and then illustrates the finite sample properties of our estimator through some Monte Carlo simulation results. Section 4 applies our method to the estimation of the black-white wage gap among males. Finally, Section 5 concludes. The appendix gathers some additional details on the data and the main proofs of our theoretical results, while the online appendix (see D'Haultfoeuille et al., 2017) presents some further theoretical results on identification and estimation, and collects the technical lemmas used in our proofs.

2 The set-up and semiparametric estimation

2.1 Model and identification

Before presenting the model, let us introduce some definitions and the notation. We denote by U_j the jth component of any given random vector $U \in \mathbb{R}^d$. For any random variable U, we denote by $\text{Supp}(U)$, F_U and S_U its support, cumulative distribution function (cdf.) and survival function, while Q_U denotes its quantile function, $Q_U(\tau) = \inf\{u : F_U(u) \geq \tau\}$. We also use some notions from extreme value theory. 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.

Let Y^* denote the outcome of interest and $X = (X_1, X_{-1}) \in \mathbb{R}^d$ denote a vector of covariates, excluding the intercept. We are interested in the average marginal effects of each component of X_1 on Y^*, with the understanding that marginal effects refer to a change from 0 to 1 for binary variables. Identification of these effects is complicated by a sample selection issue, as we only observe $(D, Y = DY^*, X)$, where D denotes the selection dummy. Moreover, 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, following D'Haultfoeuille and Maurel (2013), we rely hereafter on restrictions on the outcome equation and on an independence at infinity condition. We first suppose that the effect of the d_1-vector X_1 on Y^* is homogeneous across the distribution.

Assumption 1. *(Homogeneous effect of X_1) $Y^* = X_1'\beta_1 + \varepsilon$ where, for any $\tau \in (0,1)$, $Q_{\varepsilon|X}(\tau|X) = \beta_0(\tau) + X_{-1}'\beta_{-1}(\tau)$. $\mathcal{Q}_X = E\left[\overline{X}\,\overline{X}'\right]$ is nonsingular, with $\overline{X} = (X_1', 1, X_{-1}')'$.*

Several remarks on this assumption are in order. First, linearity is not needed for identification. Following D'Haultfoeuille and Maurel (2013), we could replace $X_1'\beta_1$ by $\psi_1(X_1)$ and $\beta_0(\tau) + X_{-1}'\beta_{-1}(\tau)$ by $\psi_2(X_{-1}, \tau)$, for any functions ψ_1 and ψ_2, without affecting our identification results. Nonetheless, imposing linearity is useful in terms of estimation as it allows us to use linear quantile regressions, resulting in computationally simple estimators, and faster convergence rates relative to a nonparametric specification. Second, Assumption 1 implies conditional independence between X_1 and ε, namely $X_1 \perp\!\!\!\perp \varepsilon | X_{-1}$. This condition is substantially weaker than the full independence assumption commonly imposed in the context of selection models (see, e.g., Chamberlain, 1986, Ahn and Powell, 1993).[5] While special regressors also rely on similar conditional independence assumptions (see Lewbel, 2014, for an overview), a crucial distinction here is that we do not restrict the support of X_1; it can be binary, for instance. Finally, and importantly, we will see below that Assumption 1 is testable.

We also need to impose some restrictions on the upper tail of ε.

Assumption 2. *(Tail and regularity of the residual)*
(i) $\sup(Supp(\varepsilon|X)) = \infty$ and $E[\exp(b \max(0, \varepsilon))] < +\infty$ for some $b > 0$.
(ii) $Supp(X)$ is compact, $x \mapsto S_{\varepsilon|X}(z|x)$ is continuous for any $z \in \mathbb{R}$ and almost surely, $\exp(\varepsilon)|X$ is in the domain of attraction of generalized extreme value distributions.
(iii) There exists $A > 0$ such that, almost surely, $S_{\varepsilon|X}(\cdot|X)$ is differentiable with increasing derivative on $[A, +\infty)$.

Assumption 2(i) is satisfied for the normal and exponential distributions, for instance. In our application to the black-white wage gap, the outcome variable Y^* is the log-wage. Given our specification, $\exp(Y^*)$, i.e. the wage, has the same tail as $\exp(\varepsilon)$ given X. As long as there exists any $\alpha > 0$ such that $E(\text{wage}^\alpha) < \infty$, Assumption 2(i) holds. It follows that this condition holds even if wages exhibit very fat tails, for instance Pareto-like. Assumption 2(ii) and (iii) are satisfied by most well-known distributions such as, again, the normal and exponential distributions. A counterexample satisfying (i) and (iii) but not (ii) is $S_{\exp(\varepsilon)|X}(t|x) = (1 + .5\sin(\ln(t))/t^2$ on $[1, +\infty]$. Such a survival function does not belong to

[5]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.

the domain of attraction of generalized extreme value distributions because of the oscillations of the sine function.

Under Assumption 1, β_1 corresponds to the average marginal effects of X_1 on Y^*. Identification of this parameter is primarily based on the following independence at infinity assumption.

Assumption 3. *(Independence at infinity) There exists $h \in (0,1]$ such that for all $x \in Supp(X)$,*

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

The main restriction in Assumption 3 is that h does not depend on x. In other words, we require selection to become independent of the covariates at infinity, that is conditional on having arbitrarily large outcomes.[6] The underlying intuition is that, if selection is 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. This condition includes as an important special case the "no selection at infinity" situation where the selection probability tends to one for large values of the outcome ($h = 1$). For instance, in the context of selection into employment, Assumption 3 holds with $h = 1$ if individuals with arbitrarily large potential wages join the workforce and are employed with a probability approaching 1. But our framework also accommodates more general forms of selection since $h < 1$ is also allowed for. In particular, $h < 1$ is needed to extend the previous example by allowing for search frictions.

We discuss in Section 1 of our supplement sufficient conditions for Assumption 3 in the context of standard threshold crossing selection models. Assumption 3 may be satisfied even if the dependence between the residuals from the selection and outcome equations is very weak. For instance, it holds for all Gaussian copulas with positive dependence. Finally, Assumption 3 fails under exogenous selection where $D \perp\!\!\!\perp Y^*|X$ and $P(D=1|X)$ is not constant. In such a case, $\lim_{y\to\infty} P(D=1|X=x, Y^*=y) = P(D=1|X=x)$ does depend on x. However, we prove in the supplementary material that our identification result (Theorem 2.1 below) still holds in this case under some additional restrictions on ε beyond Assumption 2.

Under the previous assumptions, β_1 is identified, as the following theorem shows.

Theorem 2.1. *If Assumption 1, 2(i), and 3 hold, then β_1 is identified. If Assumption 2(ii)*

[6]Given the specification of the outcome equation (see Assumption 1), the condition of independence at infinity can be equivalently rewritten as $\lim_{v\to\infty} P(D=1|X=x, \varepsilon=v) = h$.

and (iii) also hold, then, as $\tau \to 0$,

$$
\begin{aligned}
Q_{-Y|X}(\tau|X) &= Q_{-Y^*|X}(\tau/h|X) + o(1) && (2.1)\\
&= -X_1'\beta_1 - \beta_0(1-\tau/h) - X_{-1}'\beta_{-1}(1-\tau/h) + o(1). && (2.2)
\end{aligned}
$$

The first part of Theorem 2.1 is based on D'Haultfoeuille and Maurel (2013). The proof of D'Haultfoeuille and Maurel (2013), however, does not directly yield an estimation method. The second part of Theorem 2.1 is then important, as it shows that under some additional, weak conditions, β_1 can be identified using extremal conditional quantiles of Y, thus suggesting an estimation method based on extremal quantile regression. The intuition underlying this last result is as follows. For concreteness, consider the example of wages and labor market participation, and suppose that $h = 1$. Consider a high quantile q of potential wages. The probability of not participating conditional on $Y^* \geq q$ is very small, by Assumption 3. It follows that we can expect the corresponding quantile of $Y (= DY^*)$ to be close to q.[7] This is precisely what Equation (2.1) formalizes.

2.2 The estimator and its asymptotic properties

Theorem 2.1 suggests the use of extremal quantile regressions to estimate β_1. We thus define

$$\left(\widehat{\beta}_1', \widehat{\beta}_0(1-\tau_n/h), \widehat{\beta}_{-1}'(1-\tau_n/h)\right)' = \arg\min_{\beta} \sum_{i=1}^{n} \rho_{\tau_n}(-Y_i + \overline{X}_i'\beta), \qquad (2.3)$$

where $\rho_\tau(u) = (\tau - \mathbb{1}\{u < 0\})u$ is the check function used in quantile regressions and τ_n is a quantile index tending to 0. To derive the asymptotic properties of $\widehat{\beta}_1$, we rely on the asymptotic properties of extremal quantile regressions, established by Chernozhukov (2005). An important distinction though is that (2.2) has a remainder term and is therefore not exactly linear in parameters. This implies that a bias term caused by selection comes into play. An upper bound of this bias turns out to be

$$\mathcal{B}(\tau) = E\left[\sup_{t \geq 1-\tau/h} |h - P(D = 1|X, F_{Y^*|X}(Y^*|X) = t)| \times \|\overline{X}\|\right].$$

In addition to Assumptions 1–3, our asymptotic analysis relies on the three conditions below.

Assumption 4. *(i.i.d. sampling) $(D_i, Y_i, X_i)_{i=1\ldots n}$ are independent, with the same distribution as (D, Y, X).*

[7] The value 0 has no particular meaning: the true Y^* of individuals such that $D = 0$ may be lower or greater than 0. Any value $y \in \mathbb{R}$ could be imputed instead of 0, without modifying (2.2).

Assumption 5. *(Asymptotic location-scale model) There exists $\beta_{-1,r} \in \mathbb{R}^{d-d_1}$, $A > 0$, a survival function S_η, and a function H such that (i) $\inf_{x_{-1}\in Supp(X_{-1})} H(x_{-1}) > 0$; (ii) as $z \to +\infty$, uniformly in $x \in Supp(X)$,*

$$S_{U|X}(z|x) \sim S_\eta\left(\frac{z}{H(x_{-1})}\right) \tag{2.4}$$

with $U = \varepsilon - X'_{-1}\beta_{-1,r}$; (iii) S_η is differentiable with increasing derivative on $[A, +\infty)$.

Assumption 6. *(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) $\sqrt{\tau_n n}\mathcal{B}(\tau_n) \to 0$.*

Assumption 5 is the tail independence condition in Chernozhukov (2005, see condition R1 p.809) with X replaced by X_{-1}.[8] This assumption means that our model is close to a location-scale model at the upper tail.[9] Condition (iii) also imposes a mild regularity restriction on this location-scale model. But, importantly, Assumption 5 does not impose any constraint on the rest of the distribution of Y^*.

Assumption 6 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 with an intermediate order sequence τ_n, 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, $\widehat{\beta}_1$ is 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 and important feature that it guarantees asymptotic normality rather than convergence towards a non-standard, data-dependent, distribution (see Chernozhukov, 2005 and Chernozhukov and 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 (2.2) is not exactly linear vanishes quickly enough. Importantly, there always exists a τ_n satisfying Assumption 6, as Lemma 2.1 below shows. An issue is that the τ_n we exhibit in the proof depends on the term $\mathcal{B}(\cdot)$, which is unknown to the researcher. We come back to the issue of the practical choice of τ_n in Subsection 2.3.

Lemma 2.1. *Under Assumption 3, there exists $(\tau_n)_{n\geq 0}$ satisfying Assumption 6.*

[8] This is due to the fact that in our case $Q_{\varepsilon|X}(\tau|X)$ does not depend on X_1.

[9] With a location scale model $Y^* = X'\beta + (1 + X'_{-1}\delta)\varepsilon$, $\inf_{x_{-1}\in \text{Supp}(X_{-1})} 1 + x'_{-1}\delta > 0$, Conditions (i)-(ii) hold by taking $\beta_{-1,r} = \beta_{-1}$, $S_\eta = S_\varepsilon$ and $H(x_{-1}) = 1 + x'_{-1}\delta$.

The main result of this subsection is Theorem 2.2 below, which shows that the estimator of β_1 is consistent, asymptotically normal, and that the bootstrap is consistent.[10] Before stating the result, we need to introduce the following notation. Let $\mathcal{Q}_H = E\left[H^{-1}(X_{-1})\overline{X}\,\overline{X}'\right]$, $\Omega_0 = \mathcal{Q}_H^{-1}\mathcal{Q}_X\mathcal{Q}_H^{-1}$ and Ω_1 denote the first $d_1 \times d_1$ block of Ω_0. Finally, let $\lambda_n = \sqrt{n/\tau_n}hf_{-\eta}(Q_{-\eta}(\tau_n/h))$.

Theorem 2.2. *If Assumptions 1–6 hold, then*

$$\lambda_n(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} \mathcal{N}(0, \Omega_1),$$

where $\lambda_n \to \infty$. Moreover, the bootstrap is consistent for $\widehat{\beta}_1$.

Several comments on this theorem are in order. First, it shows that the rate of convergence of $\widehat{\beta}_1$ depends on τ_n, which itself depends on $\mathcal{B}(.)$. Intuitively, if $\mathcal{B}(\tau)$ tends to 0 quickly as $\tau \to 0$, Assumption 6(iii) holds even when τ_n tend to zero very slowly, implying fast rates of convergence. We formalize this intuition in Proposition 1.2 of our supplement, and provide a condition for polynomial convergence rate of $\widehat{\beta}_1$. We also show that for some data generating processes, an adequate choice of τ_n can make the rate of convergence arbitrarily close to the parametric root-n rate.

Second, and related to the first point, the convergence rate for $\widehat{\beta}_1$ is unknown. Hence, in order to conduct analytical inference, we have to estimate both the asymptotic variance and the convergence rate. On the other hand, inference by bootstrap can be conducted without further complication. Indeed, with the bootstrap, one can estimate consistently Ω_1 up to the convergence rate, which is sufficient to conduct inference on β_1. In the following, we denote by $\widehat{\Omega}$ such estimator.[11]

Finally, we can actually obtain a multivariate generalization of Theorem 2.2, where we consider (with a slight abuse of notation) several estimators $\widehat{\beta}_1(\ell_1\tau_n), ..., \widehat{\beta}_1(\ell_J\tau_n)$ of β_1, with $\ell_1 < ... < \ell_J = 1$, instead of just one. It turns out that that the best combination of these estimators, in terms of asymptotic variance, is simply $\widehat{\beta}_1(\tau_n) = \widehat{\beta}_1$.[12] In other words, there is no efficiency gain in computing several quantile regressions. On the other hand, we can use such a result

[10] Consider an estimator $\widehat{\theta}$ of a parameter θ such that for some sequence $(r_n)_n$, $r_n(\widehat{\theta} - \theta)$ converges in distribution, and let $\widehat{\theta}^*$ denote the bootstrap counterpart of $\widehat{\theta}$. We say that the bootstrap is consistent for $\widehat{\theta}$ if with probability tending to one and conditional on the sample, $r_n(\widehat{\theta}^* - \widehat{\theta})$ converges to the same distribution as $r_n(\widehat{\theta} - \theta)$. We refer to, e.g., van der Vaart and Wellner (1996), Section 3.6, for a formal definition of conditional convergence.

[11] Bootstrap estimators $\widehat{\Omega}$ of Ω_1 will satisfy $\lambda_n^2\widehat{\Omega}\Omega_1^{-1} \xrightarrow{p} I_{d+1}$. Given the bootstrap sample $\{\widehat{\beta}_1^b\}_{b=1}^B$, we can follow Machado and Parente (2005) to compute $\widehat{\Omega}$ via percentile method, which is justified theoretically. Or we can simply compute $\widehat{\Omega} = \frac{1}{B}\sum_{b=1}^B(\widehat{\beta}_1^b - \widehat{\beta}_1)(\widehat{\beta}_1^b - \widehat{\beta}_1)'$, which is easier to implement.

[12] We thank a referee for raising this point.

to build a specification test. Specifically, let us consider the J-test statistic

$$T_J(\ell) = [(1/\ell) - 1]^{-1}(\widehat{\beta}_1(\tau_n) - \widehat{\beta}_1(\ell\tau_n))'\widehat{\Omega}^{-1}(\widehat{\beta}_1(\tau_n) - \widehat{\beta}_1(\ell\tau_n)), \tag{2.5}$$

and consider the test $\varphi_\alpha(\ell) = \mathbb{1}\{T_J(\ell) > q_{d_1}(1-\alpha)\}$, where $q_{d_1}(1-\alpha)$ is the quantile of order $1-\alpha$ of a χ^2 distribution with d_1 degrees of freedom. The following theorem shows that such a test has a correct asymptotic level, and power against suitably defined local alternatives.

Theorem 2.3. *Suppose that Assumptions 2–4 and 6 hold. Then:*

1. *If Assumptions 1 and 5 also hold, then, for any* $0 < \ell < 1$, $\lim_{n\to\infty} E(\varphi_\alpha(\ell)) = \alpha$.
2. *If we consider the sequence of local alternatives defined by Assumptions 1' and 5' in the appendix,* $\lim_{n\to\infty} E(\varphi_\alpha(\ell)) > \alpha$ *and* $\ell \mapsto \lim_{n\to\infty} E(\varphi_\alpha(\ell))$ *is maximal for* $\ell^* = \arg\max_{\ell\in[0,1]} \ell[\ln(l)]^2/(1-\ell)$, $\ell^* \simeq 0.2$.

This test has similarities with a Hausman test, since it consists in comparing two estimators of β_1, one of which being efficient (among quantile estimators) under the null. A difference with the usual Hausman test, however, is that both estimators are inconsistent under the alternative. We show nonetheless that the test is consistent against local alternatives where X_1 has a scale effect on Y^* in addition to its location effect, and prove that under such alternatives the test has maximal power for $\ell = \ell^* \simeq 0.2$.[13]

2.3 Choice of the quantile index

The estimator of β_1 is asymptotically normal with zero mean provided that it is based on a sequence of quantile indices τ_n satisfying the bias-variance trade-off of Assumption 6. Though there always exists a sequence τ_n satisfying Assumption 6 under Assumption 3, admissible rates of convergence towards 0 for τ_n are unknown, since they depend on $\mathcal{B}(\tau_n)$, which is itself unknown. A related issue arises in the estimation at infinity of the intercept of sample selection models (see Andrews and Schafgans, 1998 and Schafgans and Zinde-Walsh, 2002) or in the estimation of extreme value index (see Drees and 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.

[13] Note that, by design, this test may only detect deviations from our conditional location-shift specification occurring in the tails.

Specifically, let us consider the same test statistic as before, but where $(\ell\tau_n, \tau_n)$ are replaced by $(\ell_1\tau_n, \ell_2\tau_n)$, with $\ell_1 < 1 < \ell_2$:

$$T_J(\tau) = [1/\ell_1 - 1/\ell_2]^{-1}(\widehat{\beta}_1(\ell_2\tau) - \widehat{\beta}_1(\ell_1\tau))'\widehat{\Omega}^{-1}(\widehat{\beta}_1(\ell_2\tau) - \widehat{\beta}_1(\ell_1\tau)).$$

A slight adaptation of the proof of Theorem 2.3 shows that if τ_n satisfies Assumption 6, $T_J(\tau_n)$ converges to a chi-squared distribution with d_1 degrees of freedom as n grows to infinity. In Section 3 of the supplementary material, we also show that otherwise, the asymptotic distribution of $T_J(\tau_n)$ includes an additional bias term. Heuristically, this suggests in particular that if the median of the test statistic is close enough to the median of a chi-squared distribution with d_1 degrees of freedom, denoted by M_{d_1}, then the bias term should be small. Our data-driven procedure builds on this simple 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 $\mathcal{G}$, 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

$$\widehat{\text{diff}}_n(\tau) = \frac{|M_s(\tau) - M_{d_1}|}{\sqrt{b_n\tau}},$$

where b_n denotes the subsample size.

Similarly, the asymptotic covariance matrix is estimated by the covariance matrix of the subsampling estimator of β_1, multiplied by the normalizing factor b_n/n. We sum up the diagonal elements of this matrix and call this sum $\widehat{\text{Var}}_n(\tau)$. We then select the quantile index as follows:

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

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. As in the case of nonparametric regressions, this is needed to control the asymptotic bias that would otherwise affect the limiting distribution of our estimator.

3 Implementation and simulations

3.1 Details on implementation

In this subsection, we provide a detailed algorithm for computing our estimator, which is used in our simulations and in the application. First, we have to fix some tuning parameters in our procedure on the choice of τ_n. We choose $(\ell_1, \ell_2) = (0.9, 1.1)$. We also set the subsampling size b_n equal to

$$b_n = 0.6n - 0.2(n-500)^+ - 0.2(n-1000)^+ - 0.2\left[1 - \frac{\ln(2000)}{\ln(n)}\right](n-2000)^+,$$

with $x^+ = \max(0, x)$. Finally, we have to choose a grid $\mathcal{G}$. We set the upper bound of this interval to 0.3, the lower bound to $\min(0.1,\ 80/b_n)$, and the number of (uniformly spaced) grid points equal to 40. The 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 and Fernandez-Val, 2011 for a related discussion). Our estimator can be computed through the following steps:

1. Draw B bootstrap samples and S subsamples of size b_n. In our simulations and application, we let $B = S = 150$.
2. For each $\tau \in \mathcal{G}$:
   (a) Compute the estimator of $\beta(\tau) = (\beta_1', \beta_0(1-\tau/h), \beta_{-1}'(1-\tau/h))'$:

   $$\widehat{\beta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(-Y_i + \overline{X}_i'\beta).$$

   (b) Compute

   $$\widehat{\Omega}(\tau) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\beta}^b(\tau) - \widehat{\beta}(\tau))(\widehat{\beta}^b(\tau) - \widehat{\beta}(\tau))',$$

   with $\widehat{\beta}^b(\tau)$ the bootstrap estimator of $\beta(\tau)$ on the b-th bootstrap sample.

   (c) Compute, for each subsample $s = 1 \dots S$, the estimator of β_1 ($\widehat{\beta}_1^s(\tau)$), and the J-test statistic:[14]

   $$T_J^s(\tau) = (b_n/n)[1/\ell_1 - 1/\ell_2]^{-1}(\widehat{\beta}_1^s(\ell_2\tau) - \widehat{\beta}_1^s(\ell_1\tau))'\widehat{\Omega}^{-1}(\tau)(\widehat{\beta}_1^s(\ell_2\tau) - \widehat{\beta}_1^s(\ell_1\tau)).$$

   (d) Compute $\widehat{\text{diff}}_n(\tau) = \frac{|M_S(\tau) - M_{d_1}|}{\sqrt{b_n\tau}}$ where $M_S(\tau)$ denotes the median of the $(T_J^s(\tau))_{s=1}^S$.

   (e) Compute $\widehat{\text{Var}}_n(\tau) = (b_n/n)\sum_{k=1}^{d_1}\widehat{\Sigma}(\tau)_{kk}$, where $\widehat{\Sigma}(\tau)_{kk}$ is the $k-$th diagonal term of

   $$\widehat{\Sigma}(\tau) = \frac{1}{S}\sum_{s=1}^{S}(\widehat{\beta}_1^s(\tau) - \overline{\beta}_1(\tau))(\widehat{\beta}_1^s(\tau) - \overline{\beta}_1(\tau))' \quad \text{with} \quad \overline{\beta}_1(\tau) = \frac{1}{S}\sum_{s=1}^{S}\widehat{\beta}_1^s(\tau).$$

[14] The term b_n/n accounts for the fact that the rate of convergence of the J statistic on the subsample is $\lambda_{b_n}^2$ instead of λ_n^2, with $\lambda_n^2/\lambda_{b_n}^2 = n/b_n$.

3. Compute $\widehat{\tau}_n = \arg\min_{\tau\in\mathcal{G}} \widehat{\mathrm{Var}}_n(\tau) + \widehat{\mathrm{diff}}_n(\tau)$.

4. Define $\widehat{\beta}_1 = \widehat{\beta}_1(\widehat{\tau}_n)$, $\widehat{\Omega} = \widehat{\Omega}(\widehat{\tau}_n)$. Optionally, compute $\widehat{\beta}_1(0.2\widehat{\tau}_n)$ and then $T_J(0.2)$, as defined in (2.5), to perform the specification test of the model.

The multiple steps involved in our estimation procedure are primarily due to the fact that we use bootstrap for inference and subsampling for computing $\widehat{\tau}_n$. Importantly though, the overall estimation procedure remains computationally easily tractable. For instance, the whole procedure takes less than 8 minutes (about 440 seconds) with Matlab and Stata, with the initial NLSY79 sample ($n = 1,674$) that we use in our application (see Subsection 4.1 below).[15]

3.2 Monte Carlo 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, $V(\varepsilon) = V(\eta) = 1$ and $\mathrm{Cov}(\varepsilon, \eta) = 0.2$. 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 $\delta_3 = -0.3$. Throughout this section we focus on the performances of our estimator of β_1. In our application, the black-white wage gap will be estimated similarly.

We report in Table 1 below, for each sample size, the bias, standard deviation and root-mean-square error (RMSE) of our estimator of the parameter of interest β_1 ("Extremal" column), and of a naive OLS estimator obtained using only the observations such that $D = 1$ and ignoring the selection problem.[16] Finally, we also report in this table the average quantile

[15] These CPU times are obtained on an Intel Core$^{\mathrm{TM}}$, i5-4200U, 1.6 GHz with 8Gb of RAM. Most of the CPU time is actually consumed by the computation of the data-driven $\widehat{\tau}_n$: for a given τ_n, the computational time is only around 30 seconds when including the bootstrap for inference.

[16] We also computed the finite-sample performances of the maximum likelihood estimator. As expected, the MLE is more efficient than our distribution-free estimator. Specifically, for n=2,000, the RMSE associated with the MLE is equal to 0.045, against 0.067 for our estimator (detailed results are available upon request). We also examine in the supplement the properties of our estimator under a very similar DGP, but with exogenous selection. Though not necessarily consistent in this case, the estimator actually displays very good finite sample performances, even better than the OLS for $n = 1,000$.

indices computed across all simulations.

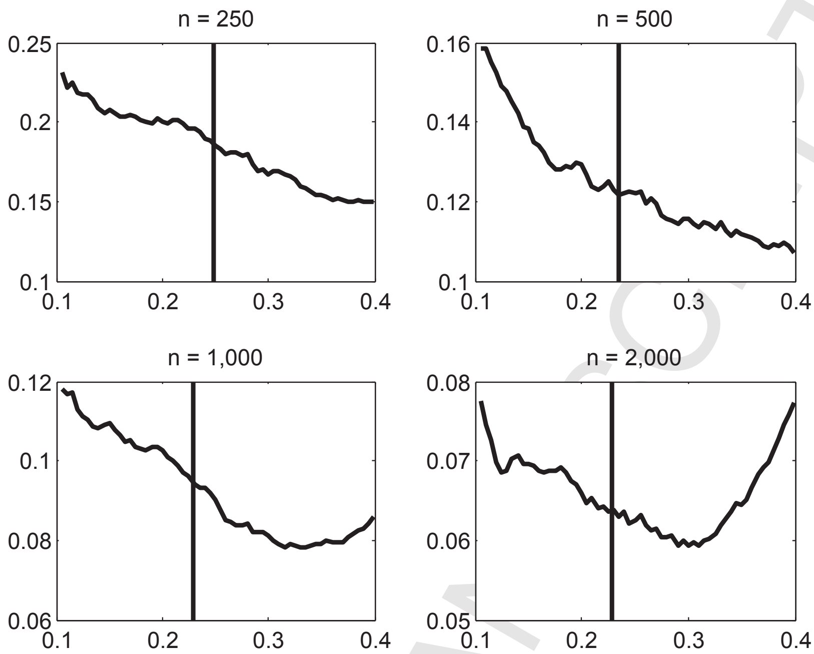
	Extremal	OLS		Extremal	OLS
$n = 250$			$n = 1,000$		
Bias	-0.018	-0.089	Bias	0.005	-0.078
Std. dev.	0.183	0.150	Std. dev.	0.094	0.075
RMSE	0.184	0.174	RMSE	0.094	0.108
Average τ_n	0.249		Average τ_n	0.230	
$n = 500$			$n = 2,000$		
Bias	-0.009	-0.077	Bias	-0.004	-0.083
Std. dev.	0.127	0.099	Std. dev.	0.067	0.048
RMSE	0.127	0.126	RMSE	0.067	0.096
Average τ_n	0.235		Average τ_n	0.228	

Note: Results were obtained using 280 simulations for each sample size.

Table 1: Monte Carlo simulations: Extremal and OLS estimator of β_1.

First, for each sample size, the bias-standard deviation ratio of our estimator is much smaller than 1, consistent with our data-driven choice of τ_n, which is aimed at undersmoothing. In practice, our estimator exhibits a fairly small bias for sample sizes larger than $n = 500$. The OLS estimator, on the other hand, displays a large and nonvanishing bias across all sample sizes. Besides, the standard deviation of our estimator decreases with the sample size, as expected given the consistency of our estimators.

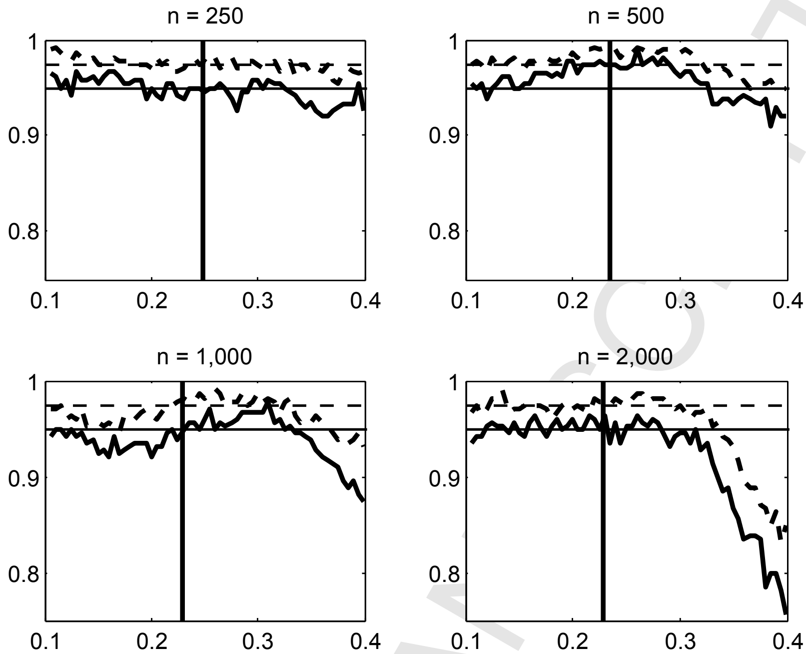
Figure 1 displays the RMSE of our estimator $\widehat{\beta}_1$ as a function of the quantile index τ_n. The plots corresponding to $n = 1,000$ and $n = 2,000$ exhibit a U-shaped relationship between the RMSE and the quantile index. This pattern reflects a bias-variance tradeoff with respect to the choice 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 RMSE and τ_n is mostly decreasing for $n = 250$ and $n = 500$. This is consistent with the variance term dominating the bias term for such small sample sizes. The vertical line corresponds to the average τ_n (across simulations) obtained with our data-driven method. For all sample sizes, this index is smaller than the one yielding the smallest RMSE, consistent with our data-driven method tending to undersmooth. However, for sample sizes larger than 250, the RMSE evaluated at the average selected quantile index is close to the minimum.



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

Figure 1: Relationship between RMSE of $\widehat{\beta}_1$ (Y-axis) and τ_n (X-axis)

Next, we examine in Figure 2 the relationship between the coverage of the 95% and 97.5% confidence intervals constructed with our estimator $\widehat{\beta}_1$ and the quantile index τ_n. The coverage is generally quite close to the nominal rates for values of τ_n in the neighborhood of the average quantile index obtained with our data-driven method, although both confidence intervals tend to be conservative for $n = 500$. The sharp decline in coverage for larger values of the quantile index for $n = 1,000$ and $n = 2,000$ reflects the existence of a nonvanishing bias for fixed values of τ_n. This stresses the importance of carefully choosing the quantile index in order to conduct valid inference on the parameters of interest. Overall, these simulation results indicate that our data-driven procedure does a good job in selecting appropriate quantile indices.



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 solid and dashed curves represent the coverages of the 95% and 97.5% confidence intervals, as a function of the quantile index τ_n. The confidence intervals are constructed using percentile bootstrap.

Figure 2: Relationship between coverage on β_1 (Y-axis) and τ_n (X-axis)

Finally, we conclude this section by examining in Table 2 the finite sample performance of our specification test based on $T_J(\ell)$, for $\ell \in \{0.2, 0.4, 0.6, 0.8\}$ and $\alpha = 0.05$. For that purpose, we consider location-scale alternatives, i.e. models for which $\delta_1 \neq 0$. Our results show that the test is generally conservative for $\delta_1 = 0$, for all values of ℓ. This could be expected since the data-driven τ_n is computed by minimizing the sum of two terms, one of which being close to $T_J(\ell)$. Nevertheless, and importantly, the test exhibits good power against alternatives, especially for $n \geq 1,000$. As n increases, we also see that the power is maximized for $\ell = 0.2$, consistent with the theory (Theorem 2.3), although the power remains large for $\ell = 0.4$.

	$\delta_1 = 0$				$\delta_1 = 0.5$				$\delta_1 = 1$			
ℓ	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
n=250	0.007	0.039	0.032	0.032	0.129	0.111	0.132	0.064	0.268	0.296	0.304	0.086
n=500	0.018	0.039	0.046	0.021	0.171	0.232	0.164	0.061	0.396	0.446	0.350	0.068
n=1,000	0.007	0.032	0.043	0.025	0.329	0.361	0.271	0.071	0.682	0.682	0.514	0.200
n=2,000	0.011	0.018	0.075	0.021	0.575	0.514	0.386	0.089	0.896	0.871	0.721	0.325

Note: Power of the test for a nominal level of $\alpha = 0.05$. The results were obtained using 280 simulations for each sample size.

Table 2: Local power of the specification test

4 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 and 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 and Welch, 1989, Neal and Johnson, 1996, Juhn, 2003, Neal, 2004, Neal, 2006 and Neal and 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 and 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.[17]

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

4.1 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 and Johnson (1996) on this question. We use the same sample as Neal and 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 and Johnson (1996) for a detailed discussion on the data.

We start by replicating the results of Neal and Johnson (1996) in Table 3 below. We run 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 and 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 and 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 3) relaxes this assumption by replacing the median with some extremal quantile of the conditional wage distribution.[18]

As is put forward by Neal and 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. The estimated black-white wage differential further 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 and Johnson (1996) (see Columns (2) and (4)).

whites.

[18] Our identifying condition is also weaker in the sense that h does not need to be equal to 1. Note however that our estimation method is not strictly less restrictive than the imputation - median regression method used by Neal and Johnson (1996). Contrary to Neal and Johnson, we need to impose a conditional location-shift model on X_1. This reflects an underlying trade-off between the strength of the identifying assumption on the selection process, and the restrictions on the (conditional) distribution of potential outcomes.

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

We now investigate how the above results change when we use our method to estimate the black-white wage gap. In the discussion below we focus on the black coefficient (or equivalently on the black-white wage gap) which is our parameter of interest here. All of these estimation results are obtained after controlling for age, AFQT, AFQT squared, and Hispanic ethnicity. In Table 4 we present the results obtained when implementing the estimator defined in Section 2.2 ("Extremal" column).[19] We compare our estimation results with the results from a median log-wage regression with zero log-wage imputation for non-participants ("Median" column), and those from a naive OLS estimator that ignores the selection issue ("OLS" column).[20]

[19] In practice, wages are likely measured with error. While the estimation method discussed in Section 2 does not explicitly allow for measurement errors, we show in Section 2.1 of the supplementary material that, in the special case of a location shift model, our framework is in fact robust to classical measurement errors affecting the outcome. We also discuss in that section the effect of measurement errors on the asymptotic variance of the estimator.

[20] A potential alternative would be to estimate the black-white wage gap using the inverse density weighting scheme of Lewbel (2007), treating AFQT as a special regressor. In this context, AFQT appears to be the only candidate as a special regressor. 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 indicates that this method cannot be used in this context.

	Extremal	Median	OLS
Black	-0.106	-0.135	-0.071
	(0.034)	(0.034)	(0.027)
J-test (p-value)	0.238	-	-

Note: our estimator and its standard error, reported in the columns "Extremal", are computed as described in Subsection 3.1, with X_1 =Black, and X_{-1} =(Hispanic, Age, AFQT, AFQT^2)'.

Table 4: Extremal quantile regression results and alternative estimation methods (NLSY79)

Several remarks are in order. First, the p-value of the specification test introduced in Section 2.2 is equal to 0.238, implying that one cannot reject our specification at any standard statistical level. Second, the estimation results from our extremal quantile method show that the size of the black-white wage gap (10.6%) is smaller than the estimated gap obtained under the imputation method proposed by Neal and Johnson (1996) (13.5%), but larger than the gap estimated using simple OLS (7.1%).[21] That our estimate of the black-white wage gap is smaller than the one obtained with the imputation method (although not significantly so) is consistent with our estimator being based on a weaker identifying assumption on the selection process. While Neal and 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. Overall, our results are in line with the key takeaway of Neal and Johnson (1996), namely that premarket factors, as measured here by AFQT, account for most of the black-white wage differential.[22]

4.2 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

[21] Neal and Johnson (1996) also estimate the black-white wage gap for higher quantiles than the median, in particular seventy-fifth and ninetieth percentiles. They also find that the black-white gap is lower for those larger quantiles.

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

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 and 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.[23] The baseline estimation results are reported in Table 5 below. The sample sizes are equal to $1,077$ and $1,123$ for the NLSY79 and NLSY97 cohorts, respectively.

	Extremal	Median	OLS	Extremal	Median	OLS
	NLSY79			NLSY97		
Black	-0.119 (0.044)	-0.145 (0.039)	-0.081 (0.035)	-0.159 (0.043)	-0.167 (0.058)	-0.097 (0.037)
J-test (p-value)	0.814	-	-	0.773	-	-

Note: Our estimator and its standard error, reported in the columns "Extremal", are computed as described in Subsection 3.1, with X_1 =Black, and X_{-1} =(Hispanic, Age, AFQT, AFQT^2)'.

Table 5: Extremal quantile regression results and alternative estimation methods (NLSY79-NLSY97)

The estimation results obtained with our method 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 11.9% to 15.9%. On the other hand, the results from the median regression of Neal and Johnson (1996) imply a smaller across-cohort increase in the black-white wage gap (from 14.5% to 16.7%). While the naive OLS results show that fully ignoring the selection issue results in a smaller estimated increase in the black-white gap (from 8.1% to 9.7%), correcting for selection appears to be more important for the black-white wage gaps in level. In particular, for the NLSY97 cohort, the estimated wage gap increases by 6.2 points after accounting for selection using our estimation method. Finally, as illustrated by the large p-values (0.814 and 0.773 for the NLSY79 and NLSY97 samples, respectively), one cannot reject the validity of our specification for either cohort.

Do these results suggest that labor market discrimination against blacks has 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? While providing a definite answer

[23] We provide more details on the data in Appendix A.

to those questions is outside of the scope of this paper, 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 6 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 and Johnson, for three different specifications. The first specification ("No premarket factor") only controls for age and the Hispanic dummy, the second specification ("AFQT") also controls for AFQT and AFQT squared, while the third specification ("Preferred") further controls for parental education and household structure. Note that, similarly to the results discussed earlier, one cannot reject our model at any standard statistical level for any of these three specifications, with the p-values ranging from 0.235 for the specification without AFQT using the NLSY97, to 0.912 for the specification with premarket factors using the NLSY79.

	Extremal	Median	Extremal	Median
	NLSY79		NLSY97	
No premarket factor	-0.329	-0.349	-0.313	-0.311
	(0.052)	(0.032)	(0.040)	(0.051)
AFQT	-0.119	-0.145	-0.159	-0.167
	(0.044)	(0.039)	(0.043)	(0.058)
Preferred	-0.106	-0.123	-0.092	-0.135
	(0.047)	(0.042)	(0.056)	(0.064)

Notes: The "preferred" specification includes AFQT, parental education and household structure. For that case, the sample is restricted to the individuals with non-missing parental education and household structure (sample size=$1,016$ for the NLSY79, $1,071$ for the NLSY97). Our estimator and its standard error, reported in the columns "Extremal", are computed as described in Subsection 3.1. The J-test p-value for the ("No premarket factor", "AFQT", "Preferred") specifications are (0.490, 0.814, 0.912) and (0.235, 0.773, 0.721), for NLSY79 and NLSY97, respectively.

Table 6: 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 remained essentially stable across the 1979 and 1997 cohorts, with a small and insignificant 1.6 points decrease over the period of interest. This result points to a lack of progress in closing the black-white wage gap 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 worth noting that this finding is consistent with the estimates obtained by Neal and 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 and 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 the estimated black-white wage gap increases over time after controlling for AFQT, Table 6 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 stable across cohorts, decreasing by 1.4 points only (from 10.6% to 9.2%) between 1990 and 2007. This provides suggestive evidence that the across-cohort increase in the wage gap conditional on AFQT is attributable to the premarket factors other than AFQT, likely reflecting a time-varying omitted variable bias based on these family environment characteristics.

5 Concluding remarks

In this paper, we develop a new inference method for semiparametric sample selection models. The key advantage of our method is that it can be used in frequent 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. This condition is typically mild provided that selection is endogenous. Building on this identification strategy, we propose a simple estimation procedure which is based on quantile regression in the tails. We establish the consistency and asymptotic normality of our estimators by extending the analysis of Chernozhukov (2005) to a setting with sample selection, and show that bootstrap is consistent. The choice of an appropriate quantile index is important in this context, and we derive a data-driven procedure for this purpose. Finally, we build an intuitive and simple-to-implement specification test for our model. Importantly for the practical usefulness of our method, we show that our estimation procedure performs well in small samples.

We then 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 important 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 key role in explaining the black-white wage gap, and that this gap has remained essentially stable over the period 1990 to 2007.

References

Ahn, H. and J. L. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*(1-2), 3–29.

Altonji, J., P. Bharadwaj, and F. Lange (2012). Changes in the characteristics of american youth: Implications for adult outcomes. *Journal of Labor Economics 30*(4), 783–828.

Altonji, J. and R. Blank (1999). Race and gender in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3. Elsevier.

Andrews, D. K. and M. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies 65*(3), 497–517.

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*(1), 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), 1–30.

Butler, R. and J. Heckman (1977). The government's impact on the labor market status of black americans: A critical review. In F. E. B. et al. (Ed.), *Equal Rights and Industrial Relations*. Madison, Wis.: Indus. Relations Res. Assoc.

Carneiro, P., J. Heckman, and D. Masterov (2005). Labor market discrimination and racial differences in premarket factors. *Journal of Law and Economics 48*(1), 1–39.

Chamberlain, G. (1986). Asymptotic efficiency in semiparametric model with censoring. *Journal of Econometrics 32*(2), 189–218.

Chen, S., G. Dahl, and S. Khan (2005). Nonparametric identification and estimation of a censored location-scale regression model. *Journal of the American Statistical Association 100*(469), 212–221.

Chen, S. and S. Khan (2003). Semiparametric estimation of a heteroskedastic sample selection model. *Econometric Theory 19*(6), 1040–1064.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*(2), 806–839.

Chernozhukov, V. and I. Fernandez-Val (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies 78*(2), 559–589.

Chernozhukov, V., I. Fernandez-Val, and T. Kaji (2016). Extremal quantile regression: An overview. In *Handbook of Quantile Regression (Forthcoming).*

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis 76*(2), 226–248.

Das, M., W. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies 70*(1), 33–58.

D'Haultfoeuille, X. and A. Maurel (2013). Another look at the identification at infinity of sample selection models. *Econometric Theory 29*(1), 213–224.

D'Haultfoeuille, X., A. Maurel, and Y. Zhang (2017). Supplement to "Extremal quantile regressions for selection models and the black-white wage gap". Working Paper.

Donald, S. (1995). Two-step estimation of heteroskedastic sample selection models. *Journal of Econometrics 65*(2), 347–380.

Drees, H. and E. Kaufmann (1998). Selection of the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications 75*(2), 149–172.

Falk, M. (1991). A note on the inverse bootstrap process for large quantiles. *Stochastic processes and their applications 38*(2), 359–363.

Gronau, R. (1974). Wage comparison–a selectivity bias. *Journal of Political Economy 82*(6), 1119–1143.

Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*(4), 679–694.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. (1990). Varieties of selection bias. *The American Economic Review 80*(2), 313–318.

Johnson, W., Y. Kitamura, and D. Neal (2000). Evaluating a simple method for estimating black-white gaps in median wages. *American Economic Review 90*(2), 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. and E. Tamer (2010). Irregular identification, support conditions and inverse weight estimation. *Econometrica 78*(6), 2021–2042.

Lewbel, A. (2007). Endogenous selection or treatment model estimation. *Journal of Econometrics 141*(2), 777–806.

Lewbel, A. (2014). An overview of the special regressor method. In J. Racine, L. Su, and A. Ullah (Eds.), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press.

Machado, J. A. and P. Parente (2005). Bootstrap estimation of covariance matrices via the percentile method. *The Econometrics Journal 8*(1), 70–78.

Neal, D. (2004). The measured black-white wage gap among women is too small. *Journal of Political Economy 112*(1), 1–28.

Neal, D. (2006). Why has black-white skill convergence stopped? In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1. Elsevier.

Neal, D. and W. Johnson (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy 104*(5), 869–895.

Neal, D. and A. Rick (2014). The prison boom and the lack of black progress after Smith and Welch. NBER Working paper 20283.

Newey, W. (2009). Two step series estimation of sample selection models. *The Econometrics Journal 12*(1), 217–229.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

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

Schafgans, M. and V. Zinde-Walsh (2002). On intercept estimation in the sample selection model. *Econometric Theory 18*(1), 40–50.

Smith, J. and F. Welch (1989). Black economic progress after Myrdal. *Journal of Economic Literature 27*(2), 519–564.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources 33*(1), 127–169.

A 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 7 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 8 reports the labor force participation rates for the NLSY79 and NLSY97 samples, separately for blacks and whites.

	NLSY79		NLSY97	
	Blacks	Whites	Blacks	Whites
AFQT	-0.716	0.387	-0.726	0.373
Std.dev.	(0.812)	(0.966)	(1.037)	(0.923)
Highest grade completed	11.638	12.859	11.239	12.849
Std.dev.	(3.927)	(3.691)	(4.811)	(4.726)
Mother high school graduate	0.447	0.715	0.707	0.829
Father high school graduate	0.368	0.665	0.518	0.758
Mother college graduate	0.040	0.093	0.107	0.217
Father college graduate	0.046	0.188	0.086	0.245
Both parents at age 14	0.486	0.760	0.264	0.597

Note: Samples restricted to males. Blacks account for 31% (25%) of the NLSY79 (NLSY97) sample.

Table 7: 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 8: Labor force participation rates (males)

B Proofs of the results

B.1 Proof of Theorem 2.1

By Assumption 1, $\varepsilon \perp\!\!\!\perp X_1|X_{-1}$. Then, conditional on X_{-1}, $Y^* = X_1'\beta_1 + \varepsilon$ is a location model. Given that $E[\exp(b\max(\varepsilon,0))] < +\infty$, $E[\exp(b\max(\varepsilon,0))|X_{-1}] < +\infty$ almost surely. Therefore, the identification result in D'Haultfoeuille and Maurel (2013) directly applies. This means that $X_1'\beta_1$ is identified up to an additive constant, c_1 say. Now, suppose that $X_1'\beta_1 + c_1 = X_1'\widetilde{\beta}_1 + \widetilde{c}_1$. Because $E\left[\overline{XX}'\right]$ is nonsingular, $\beta_1 = \widetilde{\beta}_1$ and $c_1 = \widetilde{c}_1$. Hence, β_1 is identified.

To prove (2.1), let $-z = Q_{-Y|X}(\tau|x)$ for some fixed $x \in \text{Supp}(X)$ and τ sufficiently small. By Assumption 3, there exists a function $\delta(\cdot)$ possibly depends on x, such that, for any $y \geq z$, $\delta(\tau) \downarrow 0$ as $\tau \downarrow 0$ and

$$P(D = 1|X = x, Y^* = y) \in \left(h(1+\delta(\tau))^{-1}, h(1-\delta(\tau))^{-1}\right).$$

In addition, from Assumptions 1–3 and

$$\begin{aligned} P(Y \geq z|X = x) =& P(Y^* \geq z|X = x)P(D = 1|X = x, Y^* \geq z) \\ =& \int_z^\infty P(D = 1|X = x, Y^* = y) f_{Y^*|X}(y|x)dy, \end{aligned}$$

we have that $P(Y \geq z|X = x)$ is continuous in z, $P(Y \geq z|X = x)$ has a positive derivative for any $z \in [A, \infty)$, and

$$\frac{F_{-Y^*|X}(-z|x)}{F_{-Y|X}(-z|x)} \in \left(h^{-1}(1-\delta(\tau)), h^{-1}(1+\delta(\tau))\right).$$

Therefore, $F_{-Y|X}(-z|x) = \tau$ and

$$Q_{-Y|X}(\tau|x) \in \Big(Q_{-Y^*|X}(\tau(1-\delta(\tau))/h|x), Q_{-Y^*|X}(\tau(1+\delta(\tau))/h|x)\Big).$$

Let $s(\tau) = Q_{-Y|X}(\tau|x) - Q_{-Y^*|X}(\tau/h|x)$. Then,

$$s(\tau) \in \Big(Q_{-Y^*|X}(\tau(1-\delta(\tau))/h|x) - Q_{-Y^*|X}(\tau/h|x), Q_{-Y^*|X}(\tau(1+\delta(\tau))/h|x) - Q_{-Y^*|X}(\tau/h|x)\Big).$$

We now show that the lower bound above tends to zero. By a similar reasoning, the upper bound tends to zero. The result then follows by combining $s(\tau) = o(1)$ with Assumption 1.

By the mean value theorem, there exists $\tilde{m} \in (1-\delta(\tau), 1)$ such that

$$\begin{aligned}
&Q_{-Y^*|X}(\tau(1-\delta(\tau))/h|x) - Q_{-Y^*|X}(\tau/h|x) \\
=&\int_{1/h}^{(1-\delta(\tau))/h} \frac{b(\tau)\tau dm}{f_{-Y^*|X}(Q_{-Y^*|X}(m\tau|x)|x)b(\tau)} \\
=&\Big(-\delta(\tau)b(\tau)/h\Big)\frac{\tau}{f_{-Y^*|X}(Q_{-Y^*|X}(\tilde{m}\tau|x)|x)b(\tau)}.
\end{aligned}$$

where $b(\tau) = Q_{-Y^*|X}(e\tau|x) - Q_{-Y^*|X}(\tau|x) = Q_{-\varepsilon|X}(e\tau|x) - Q_{-\varepsilon|X}(\tau|x)$.

By Exercise 1.2.6.(a) in Resnick (1987), and Assumption 2, $\varepsilon|X$ is in the attraction domain of generalized extreme value distribution. Moreover, by Lemma 5.1 of our supplement, $S_{\varepsilon|X}$ is rapidly varying at $+\infty$. Hence, $\varepsilon|X$ actually belongs to the the attraction domain of type I extreme value distribution. Therefore, by the same argument as when proving (5.3) in our supplement, we have, locally uniformly in t and as $\tau \to 0$,

$$\frac{\tau}{f_{-Y^*|X}(Q_{-Y^*|X}(t\tau|x)|x)} \sim b(t\tau).$$

Notice that , $\tilde{m} \to 1$ and $b(t\tau)/b(\tau) \to \ln(et) - \ln(t) = 1$ by (5.8) in our supplement. Hence,

$$\frac{\tau}{f_{-Y^*|X}(Q_{-Y^*|X}(\tilde{m}\tau|x)|x)b(\tau)} \to 1.$$

In addition, $\delta(\tau) \downarrow 0$ and $b(\tau)$ is bounded as $\tau \to 0$ by Lemma 5.1 in our supplement. Therefore, the lower bound of $s(\tau)$ tends to 0 and (2.1).

Finally, (2.2) follows directly from (2.1) and Assumption 1.

B.2 Proof of Lemma 2.1

First, if $\mathcal{B}(\cdot) = 0$ on $(0, c)$ for some $1 > c > 0$, then (iii) holds trivially. Otherwise define, for any $\alpha \in (0,1)$, $G(\tau) = \tau\mathcal{B}^{2(1-\alpha)}(\tau)$. By construction, $\mathcal{B}(\cdot)$ and thus $G(\cdot)$ are increasing. Then define, for n sufficiently large,

$$\tau_n^* = G^{-1}(1/n).$$

Now, for any $x \in \text{Supp}(X)$, $\lim_{\tau\to 0}\sup_{t\geq 1-\tau/h}|h - P(D=1|X=x, F_{Y^*|X}(Y^*|x)=t)| = 0$ by Assumption 3. Because this term is bounded by 2, $\lim_{\tau\to 0}\mathcal{B}(\tau) = 0$ by the dominated convergence theorem. Thus, $\lim_{\tau\to 0} G(\tau) = 0$. This implies that $\lim_{n\to\infty} G^{-1}(1/n) = 0$, ensuring that τ_n^* satisfies Condition (i). Moreover, it follows from the equality $G(\tau) = \tau\mathcal{B}^{2(1-\alpha)}(\tau)$ that

$$\frac{1}{n} = \tau_n^*\mathcal{B}^{2(1-\alpha)}(\tau_n^*),$$

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

$$\sqrt{n\tau_n^*}\mathcal{B}(\tau_n^*) = \mathcal{B}^{\alpha}(\tau_n^*) \to 0,$$

so that Condition (iii) is also satisfied.

B.3 Proof of Theorem 2.2

We decompose the proof in three steps. The first step establishes that $\lambda_n \to \infty$. The second step shows the asymptotic normality of $\lambda_n(\widehat{\beta}_1 - \beta_1)$. Finally, the third step shows the consistency of the bootstrap.

1. $\lambda_n \to \infty$

By Lemma 5.2 in our supplement, we have, for any $x = (x_1', x_{-1}')' \in \text{Supp}(X)$,

$$\begin{aligned}\lambda_n \sim \frac{\sqrt{n\tau_n}}{Q_{-\eta}(e\tau_n/h) - Q_{-\eta}(\tau_n/h)} &\sim \frac{\sqrt{n\tau_n}H(x_{-1})}{Q_{-U|X}(e\tau_n/h|x) - Q_{-U|X}(\tau_n/h|x)} \\ &= \frac{\sqrt{n\tau_n}H(x_{-1})}{Q_{-\varepsilon|X}(e\tau_n/h|x) - Q_{-\varepsilon|X}(\tau_n/h|x)}.\end{aligned}$$

By Lemma 5.1 in our supplement, $Q_{-\varepsilon|X}(e\tau_n/h|x) - Q_{-\varepsilon|X}(\tau_n/h|x)$ is bounded. Besides, $n\tau_n \to \infty$ by Assumption 6(ii). Therefore, $\lambda_n \to \infty$.

2. Asymptotic normality of $\lambda_n(\widehat{\beta}_1 - \beta_1)$

Let $\beta(\tau) = (\beta_1', \beta_0(1-\tau/h), \beta_{-1}'(1-\tau/h))'$, $\widehat{\beta}(\tau) = \left(\widehat{\beta}_1', \widehat{\beta}_0(1-\tau/h), \widehat{\beta}_{-1}'(1-\tau/h)\right)'$ and $\widehat{Z}_n = \lambda_n(\widehat{\beta}(\tau_n) - \beta(\tau_n))$. Similarly to Chernozhukov (2005, Equation (9.43)), $\widehat{Z}_n$ minimizes

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

with, for any τ,

$$W_n(\tau) = \frac{-1}{\sqrt{\tau n}}\sum_{i=1}^{n}(\tau - \mathbb{1}\{Y_i \geq \overline{X}_i'\beta(\tau)\})\overline{X}_i$$

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

$$\Lambda_n(z,\tau) = \frac{\lambda_n}{\sqrt{\tau n}} \sum_{i=1}^{n} \int_0^{(z_1+X_i'z_2)/\lambda_n} \left[\mathbb{1}\{-Y_i + \overline{X}_i'\beta(\tau) \le s\} - \mathbb{1}\{-Y_i + \overline{X}_i'\beta(\tau) \le 0\} \right] ds. \quad \text{(B.1)}$$

$\Lambda_n(z,\tau_n)$ is convex in z because the integrands are increasing in s. Moreover, by Lemma 5.4 in the supplement, $\Lambda_n(z,\tau_n) \xrightarrow{p} \frac{1}{2} z'\mathcal{Q}_H z$. We shall now prove that

$$W_n(\tau_n) \xrightarrow{d} \mathcal{N}(0, \mathcal{Q}_X). \quad \text{(B.2)}$$

For that purpose, let $M_{n,i}(\tau) = \frac{-1}{\sqrt{\tau}}(\tau - \mathbb{1}\{-Y_i + \overline{X}_i'\beta(\tau) \le 0\})\overline{X}_i + \sqrt{\tau}\mu(\tau)$, with

$$\begin{aligned}\mu(\tau) =& \frac{E\left[(\tau - \mathbb{1}\{\overline{X}'\beta(\tau) \le Y\})\overline{X}\right]}{\tau} \\ =& \frac{E\left[\left(\tau - P\big(Y \ge Q_{Y^*}(1-\tau/h|X)\big|X\big)\right)\overline{X}\right]}{\tau}.\end{aligned}$$

Then

$$W_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{n,i}(\tau) - \sqrt{n\tau}\mu(\tau). \quad \text{(B.3)}$$

By Lemma 9.6 of Chernozhukov (2005), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{n,i}(\tau_n) \xrightarrow{d} \mathcal{N}(0, \mathcal{Q}_X). \quad \text{(B.4)}$$

In order to bound $\mu(\tau_n)$, we first show that, for τ sufficiently small,

$$P\big(Y \ge Q_{Y^*|X}(1-\tau/h|X), D = 0\big|X\big) = 0.$$

Note that

$$\begin{aligned} P\big(Y \ge Q_{Y^*|X}(1-\tau/h|X), D = 0\big|X\big) &\le \mathbb{1}\{Q_{Y^*|X}(1-\tau/h|X) \le 0\} \\ &= \mathbb{1}\{P(-U \le X_1'\beta_1 + X_{-1}'\beta_{-1,r}|X) \le \tau/h\} \\ &\le \mathbb{1}\{P(-U \le M|X) \le \tau/h\} \quad \text{(B.5)} \end{aligned}$$

where $M = \sup_{x \in \text{Supp}(X)} |x_1'\beta_1 + x_{-1}'\beta_{-1,r}|$. Suppose that for any small τ, we can find a corresponding x such that Since $P(U \le M|X = x) \le \tau/h$. Then we can find a convergent sequences x_n with limit x_0 (because $\mathrm{Supp}(X)$ is compact) such that

$$P(U \leq M|X = x_0) = \lim_{n\to\infty} P(U \leq M|X = x_n) = 0.$$

This contradicts the fact that the upper end-point of $U|X$ is $+\infty$. So there exists a τ_0 such that for all $x \in \mathrm{Supp}(X)$, $P(U \leq M|X = x) > \tau_0/h$. Thus, the RHS of (B.5) equals zero for $\tau \leq \tau_0$.

Therefore, for τ sufficiently small,

$$\begin{aligned}
\|\mu(\tau)\| &= \left\| \frac{E\left[\left(\tau - P\big(Y^* \geq Q_{Y^*}(1-\tau/h|X), D = 1\big|X\big)\right)\overline{X}\right]}{\tau} \right\| \\
&= \frac{1}{\tau}\int_{1-\tau/h}^{1} E|h - P(D = 1|X, F_{Y^*|X}(Y^*|X) = t)| \times \left\|\overline{X}\right\| dt \\
&\leq \mathcal{B}(\tau)/h.
\end{aligned}$$

Therefore, by Assumption 6,

$$\sqrt{n\tau_n}\mu(\tau_n) \leq \sqrt{n\tau_n}\mathcal{B}(\tau_n)/h = o(1). \tag{B.6}$$

Combined with (B.3), (B.4), and (B.6), we obtain (B.2).

Finally, by applying the convexity lemma and the same arguments as in the end of the proof of Theorem 1 in Pollard (1991), we have

$$\widehat{Z}_n = -\mathcal{Q}_H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{n,i}(\tau_n) + o_P(1) \xrightarrow{d} \mathcal{N}(0, \Omega_0). \tag{B.7}$$

The result follows since $\lambda_n(\widehat{\beta}_1 - \beta_1)$ is the subvector of the first d_1 components of $\widehat{Z}_n$.

3. Consistency of the bootstrap

Hereafter, all bootstrap counterparts are starred. Let $\{I_{n,j}\}_{j\geq 1}$ denote an i.i.d. sequence distributed as multinomial with parameter 1 and probability $(\frac{1}{n}, \cdots, \frac{1}{n})$, so that the bootstrap weight for individual i, $w_{n,i}$, satisfies $w_{n,i} = \sum_{j=1}^{n} 1\{I_{n,j} = i\}$. We show that $\widehat{Z}_n^*$ can be linearized. First, observe that, as in (B.2), $\widehat{Z}_n^*$ minimizes

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

By part 2 of Lemma 5.4 in our supplement,

$$\Psi_n^*(z,\tau_n) = W_n^*(\tau_n)'z + \frac{1}{2}z'\mathcal{Q}_H z + o_P(1).$$

Then by applying the same argument in the proof of theorem 1 in Pollard (1991), we obtain

$$\widehat{Z}_n^* = \mathcal{Q}_H^{-1}\frac{1}{\sqrt{\tau_n n}}\sum_{i=1}^n w_{n,i}(\tau_n - \mathbb{1}\{-Y_i+\gamma(1-\tau_n/h)+X_i'\beta(1-\tau_n/h)\leq 0\})\overline{X}_i + o_P(1).$$

Since $E(w_{n,i})=1$, we have

$$\left|E\left[\frac{1}{\sqrt{\tau_n n}}\sum_{i=1}^n w_{n,i}(\tau_n - \mathbb{1}\{-Y_i+\gamma(1-\tau_n/h)+X_i'\beta(1-\tau_n/h)\leq 0\})\overline{X}_i\right]\right| \leq \sqrt{\tau_n n}\mathcal{B}(\tau_n)/h \to 0,$$

which implies that

$$\widehat{Z}_n^* = (\mathcal{Q}_H)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{n,i}M_{n,i}(\tau_n) + o_P(1).$$

Now, we use the same Poisson approximation idea as in, e.g., Chapter 3.6 of van der Vaart and Wellner (1996). Let N_n be a Poisson random variable with mean n, independent of the data and of the $\{I_{n,j}\}_{j\geq 1}$. Let also $w_{N_n,i} = \sum_{j=1}^{N_n}1\{I_{n,j}=i\}$, so that $\{w_{N_n,i}\}_{i=1}^n$ are i.i.d. Poisson random variable with unit mean. The idea is to approximate $\frac{1}{\sqrt{n}}\sum_{i=1}^n(w_{n,i}-1)M_{n,i}(\tau_n)$ by $\frac{1}{\sqrt{N_n}}\sum_{i=1}^n(w_{N_n,i}-1)M_{n,i}(\tau_n)$, and then apply the central limit theorem to the latter. First consider the approximation. Let $\mathcal{I}_j = \{i : |w_{N_n,i}-w_{n,i}|\geq j\}$ and $n_j = \#\mathcal{I}_j$. Then, for any $\delta>0$, for n sufficiently large, with probability at least $1-\delta$ (see van der Vaart and Wellner, 1996, p.348),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(w_{N_n,i}-w_{n,i})M_{n,i}(\tau_n) = \text{sign}(N_n-n)\sum_{j=1}^2\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_j}M_{n,i}(\tau_n), \tag{B.8}$$

with the convention that $\sum_{i\in\mathcal{I}_j}M_{n,i}(\tau_n)=0$ when $n_j=0$. Let us show that

$$\sum_{i\in\mathcal{I}_j}M_{n,i}(\tau_n)/\sqrt{n} = o_P(1).$$

First, observe that

$$E\left[\sum_{i\in\mathcal{I}_j}M_{n,i}(\tau_n)|(I_{n,j})_{j\geq 1},N_n\right] = 0. \tag{B.9}$$

Besides, because $V(M_{n,i}(\tau_n))\to\mathcal{Q}_X$, which is bounded, we have, for n sufficiently large, and some constant $C > 0$,

$$\left| V\left[\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_j} M_{n,i}(\tau_n)|(I_{n,j})_{j\geq 1}, N_n\right]\right| \leq \frac{Cn_j}{n} \leq \frac{C|N_n - n|}{n}.$$

Thus, using a decomposition of variance, Equation (B.9), and Jensen's inequality, we get

$$\left| V\left[\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_j} M_{n,i}(\tau_n)\right]\right| \leq \frac{C\sqrt{V(N_n)}}{n} = \frac{1}{\sqrt{n}}.$$

This implies that $\sum_{i\in\mathcal{I}_j} M_{n,i}(\tau_n)/\sqrt{n} = o_P(1)$. Thus, in view of (B.8),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_{N_n,i} - w_{n,i})M_{n,i}(\tau_n) = o_P(1).$$

As a result,

$$\widehat{Z}_n^* - \widehat{Z}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_{N_n,i} - 1)M_{n,i}(\tau_n) + o_P(1).$$

Because the $\{w_{N_n,i} - 1\}_{i=1}^n$ are i.i.d., independent of the data, and satisfy $E(w_{N_n,i} - 1) = 0$ and $V(w_{N_n,i} - 1) = 1$, we obtain, conditional on the data and with probability approaching one (see, e.g., Lemma 2.9.5 of van der Vaart and Wellner, 1996),

$$\widehat{Z}_n^* - \widehat{Z}_n \xrightarrow{d} \mathcal{N}(0, \Omega_0).$$

This establishes the validity of the bootstrap for $\widehat{\beta}_1$, since $\lambda_n(\widehat{\beta}_1^* - \widehat{\beta})$ is the subvector of the first d_1 components of $\widehat{Z}_n^* - \widehat{Z}_n$.

B.4 Proof of Theorem 2.3

1. Asymptotic level

Let $\lambda_n(\ell) = \sqrt{n}hf_{-\eta}(Q_{-\eta}(\ell\tau_n/h))/\sqrt{\ell\tau_n}$. Then, similarly to the above argument, we have

$$\lambda_n(\ell)(\widehat{\beta}(\ell\tau_n) - \beta(\ell\tau_n)) = -\mathcal{Q}_H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{n,i}(\ell\tau_n) + o_P(1).$$

Also, by Lemma 5.2 in our supplement,

$$\frac{Q'_{-\eta}(\tau_n/h)}{\sqrt{\ell}Q'_{-\eta}(\ell\tau_n/h)} \sim \frac{h(Q_{-\eta}(e\tau_n/h) - Q_{-\eta}(\tau_n/h))}{\tau_n} \times \frac{\ell\tau_n}{\sqrt{\ell}h(Q_{-\eta}(el\tau_n/h) - Q_{-\eta}(\ell\tau_n/h))} \sim \sqrt{\ell}.$$

Thus, $\lambda_n(\ell) \sim \sqrt{\ell}\lambda_n$, implying that

$$\lambda_n(\widehat{\beta}(\ell\tau_n) - \beta(\ell\tau_n)) = -\mathcal{Q}_H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{M_{n,i}(\ell\tau_n)}{\sqrt{\ell}} + o_P(1). \tag{B.10}$$

Reasoning as in Step 2 of the previous proof and using the Cramer-Wold device, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{M'_{n,i}(\ell\tau_n)}{\sqrt{\ell}}, M'_{n,i}(\tau_n)\right)' \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} 1/\ell & 1 \\ 1 & 1\end{pmatrix} \otimes \mathcal{Q}_X\right).$$

We combine this with (B.7) and (B.10) to get

$$\lambda_n(\widehat{\beta}(\ell\tau_n) - \widehat{\beta}(\tau_n)) \xrightarrow{d} \mathcal{N}\left(0, [1/\ell - 1]\Omega_1\right),$$

which yields the first result.

2. Power under local alternatives

The second statement is based on the following assumptions.

Assumption 1.' $Q_{Y^*|X}(\tau|X) = X_1'\beta_1(\tau) + \beta_0(\tau) + X_{-1}'\beta_{-1}(\tau)$. $\mathcal{Q}_X = E\left[\overline{X}\overline{X}'\right]$ *is nonsingular, with* $\overline{X} = (X_1', 1, X_{-1}')'$.

Assumption 5.' *(Asymptotic location-scale model) There exists* $(b_1, \beta_{-1,r}, \gamma_n) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d-d_1} \times \mathbb{R}^{d_1}$, $A > 0$, *a survival function* S_η, *and a function* H *such that (i)* $\inf_{x_{-1}\in Supp(X_{-1})} x_1'\gamma_n + H(x_{-1}) > 0$; *(ii) for any sequence* $z_n \to +\infty$ *as* $n \to \infty$, *uniformly in* $x \in Supp(X)$,

$$S_\eta\left(\frac{z_n}{H(x_{-1}) + x_1'\gamma_n}\right) = S_{U|X}(z_n|x)(1 + o(1/\sqrt{n\tau_n})) \tag{B.11}$$

with $U = Y^* - X_1'b_1 - X_{-1}'\beta_{-1,r}$ *and* $\gamma_n\sqrt{n\tau_n} \to c$ *for some* $c \neq 0$ *in* $\mathbb{R}^{d_1}$; *(iii)* S_η *is differentiable with increasing derivative on* $[A, +\infty)$.

Note that in this context, ε in Assumption 2 is defined by $\varepsilon = Y^* - X_1'b_1$. Now, to prove the result, note first that Step 2 in the proof of Theorem 2.2 is valid even when β_1 depends on τ_n. In addition, as $n \to \infty$, $\gamma_n \to 0$ and $H(x_{-1}) + x_1'\gamma_n \to H(x_{-1})$ uniformly in x. Therefore, we have, for any $\ell \in (0, 1]$,

$$\lambda_n(\ell)(\widehat{\beta}_1(\ell\tau_n) - \beta_1(\ell\tau_n)) = -\mathcal{Q}_H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{n,i}(\ell\tau_n) + o_P(1). \tag{B.12}$$

Next, we aim to show that, for any $\ell \in (0, 1]$,

$$\beta_1(\ell\tau_n) = b_1 + \gamma_n Q_{-\eta}(\ell\tau_n/h) + o(\lambda_n^{-1}). \tag{B.13}$$

Without loss of generality, let us focus on the case $\ell = 1$. By Assumption 5(ii),

$$Q_{-Y^*|X}(\tau_n/h|X) = X_1' b_1 + X_{-1}' \beta_{-1,r} + Q_{-U|X}(\tau_n/h|X). \tag{B.14}$$

Besides, Equation (5.2) in the supplement holds when $H(x_{-1})$ is replaced by $H(x_{-1}) + x_1'\gamma_n$. This implies that uniformly in $x \in \text{Supp}(X)$,

$$(H(x_{-1}) + x_1'\gamma_n)a(\tau_n) \sim Q_{-U|X}(e\tau_n|x) - Q_{-U|X}(\tau_n|x), \tag{B.15}$$

where $a(\tau_n) = Q_{-\eta}(e\tau_n) - Q_{-\eta}(\tau_n)$. The proof of Theorem 2.1 establishes that $Q_{-U|X}(e\tau_n|x) - Q_{-U|X}(\tau_n|x)$ is bounded. Hence, (B.15) and $\inf_x H(x_{-1}) + x_1'\gamma_n > 0$ imply that $a(\tau_n)$ is bounded as well.

In addition, Equation (5.5) in the supplement also holds when $H(x_{-1})$ is replaced by $H(x_{-1}) + x_1'\gamma_n$. Hence, uniformly in $x \in \text{Supp}(X)$,

$$\frac{Q_{-U|X}(\tau_n|x) - Q_{-\eta}(\tau_n)(H(x_{-1}) + x_1'\gamma_n)}{a(\tau_n)} \to 0.$$

Together with $\inf_x H(x_{-1}) + x_1'\gamma_n > 0$, this proves that

$$s_n = \frac{Q_{-U|X}(\tau_n|x) - Q_{-\eta}(\tau_n)(H(x_{-1}) + x_1'\gamma_n)}{(H(x_{-1}) + x_1'\gamma_n)a(\tau_n)} \to 0.$$

In addition, by letting $z_n = -Q_{-U|X}(\tau_n|x)$ in (B.11), we have

$$\frac{F_{-\eta}(Q_{-\eta}(\tau_n) + s_n a(\tau_n))}{\tau_n} - 1 = o((\tau_n n)^{-1/2}).$$

Besides, we have, for some $\tilde{V} \in (0, s_n a(\tau_n))$,

$$\begin{aligned} \frac{F_{-\eta}(Q_{-\eta}(\tau_n) + s_n a(\tau_n))}{\tau_n} - 1 &= \frac{F_{-\eta}(Q_{-\eta}(\tau_n) + s_n a(\tau_n)) - F_{-\eta}(Q_{-\eta})}{\tau_n} \\ &= \frac{f_{-\eta}(Q_{-\eta}(\tau_n) + \tilde{V})s_n a(\tau_n)}{\tau_n} \\ &\sim \frac{f_{-\eta}(Q_{-\eta}(\tau_n))s_n a(\tau_n)}{\tau_n} \\ &\sim s_n, \end{aligned}$$

where the second line is by the mean-value theorem, the fourth line holds since $s_n a_n(\tau_n) \to 0$ and Equation (9.57) of Chernozhukov (2005), and the last line is by (5.3) in our supplement. This result, Equation (5.3) again and the definition of λ_n imply

$$\begin{aligned} Q_{-U|X}(\tau_n|x) - Q_{-\eta}(\tau_n)(H(x_{-1}) + x_1'\gamma_n) &= (H(x_{-1}) + x_1'\gamma_n)a(\tau_n)s_n \\ &= o(a(\tau_n)(n\tau_n)^{-1/2}) \\ &= o(\lambda_n^{-1}). \end{aligned} \quad \text{(B.16)}$$

Since both (B.14) and (B.16) hold for all x, we obtain (B.13), by Assumption 1'. As a result, using the definition of λ_n, (5.3) and (5.7), we obtain

$$\begin{aligned} \lambda_n(\beta_1(\ell\tau_n) - \beta_1(\tau_n)) &= \lambda_n\gamma_n[Q_{-\eta}(\ell\tau_n/h) - Q_{-\eta}(\tau_n/h)] + o(1) \\ &= \frac{hf_{-\eta}(Q_{-\eta}(\tau_n/h))(Q_{-\eta}(\ell\tau_n/h) - Q_{-\eta}(\tau_n/h))}{\tau_n}c + o(1) \\ &\sim \frac{h(Q_{-\eta}(\ell\tau_n/h) - Q_{-\eta}(\tau_n/h))}{(Q_{-\eta}(e\tau_n/h) - Q_{-\eta}(\tau_n/h))}c \\ &\to h\ln(\ell)c. \end{aligned} \quad \text{(B.17)}$$

Combining (B.12), (B.13) and (B.17), we finally get

$$[(1/\ell - 1)\Omega_1]^{-1/2}\lambda_n\left(\widehat{\beta}_1(\ell\tau_n) - \widehat{\beta}_1(\tau_n)\right) \xrightarrow{d} \mathcal{N}(hg(\ell)\Omega_1^{-1/2}c, I_{d_1}),$$

where I_{d_1} is the $d_1 \times d_1$ identity matrix and $g(\ell) = \left[(\ell\ln(\ell)^2)/(1-\ell)\right]^{1/2}$. Hence, the test statistic $T_J(\ell)$ is a noncentral chi-squared distribution with d_1 degrees of freedom and a noncentrality parameter equal to $[hg(\ell)]^2 c'\Omega_1^{-1}c$. Maximizing in ℓ the local power of the test is then equivalent to maximizing $g(\ell)$. The result follows.