

# INFERENCE FOR EXTREMAL CONDITIONAL QUANTILE MODELS, WITH AN APPLICATION TO MARKET AND BIRTHWEIGHT RISKS

VICTOR CHERNOZHUKOV<sup>†</sup> IVÁN FERNÁNDEZ-VAL<sup>§</sup>

**ABSTRACT.** Quantile regression is an increasingly important empirical tool in economics and other sciences for analyzing the impact of a set of regressors on the conditional distribution of an outcome. Extremal quantile regression, or quantile regression applied to the tails, is of interest in many economic and financial applications, such as conditional value-at-risk, production efficiency, and adjustment bands in (S,s) models. In this paper we provide feasible inference tools for extremal conditional quantile models that rely upon extreme value approximations to the distribution of self-normalized quantile regression statistics. The methods are simple to implement and can be of independent interest even in the non-regression case. We illustrate the results with two empirical examples analyzing extreme fluctuations of a stock return and extremely low percentiles of live infants' birthweights in the range between 250 and 1500 grams.

**KEY WORDS:** QUANTILE REGRESSION, FEASIBLE INFERENCE, EXTREME VALUE THEORY

**JEL:** C13, C14, C21, C41, C51, C53

MONTE-CARLO PROGRAMS AND SOFTWARE ARE AVAILABLE AT [WWW.MIT.EDU/VCHERN](http://WWW.MIT.EDU/VCHERN)

---

First version: May, 2002. This version: December 26, 2009. This is a revision of the paper with the original title “Extreme Value Inference for Quantile Regression.” We would like to thank very much Takeshi Amemiya, Brigham Fradsen, Hide Ichimura, Jerry Hausman, Peter Hinrichs, Keith Knight, Roger Koenker, Joseph Romano, the editor Bernard Salanie, and two referees.

<sup>†</sup> Massachusetts Institute of Technology, Department of Economics and Operations Research Center, University College London, CEMMAP. E-mail: [vchern@mit.edu](mailto:vchern@mit.edu).

<sup>§</sup> Boston University, Department of Economics. E-mail: [ivanf@bu.edu](mailto:ivanf@bu.edu).

## 1. INTRODUCTION AND MOTIVATION

Quantile regression (QR) is an increasingly important empirical tool in economics and other sciences for analyzing the impact of a set of regressors  $X$  on features of the conditional distribution of an outcome  $Y$  (see Koenker, 2005). In many applications the features of interest are the extremal or tail quantiles of the conditional distribution. This paper provides practical tools for performing inference on these features using extremal QR and extreme value theory. The key problem we address is that conventional inference methods for QR, based on the normal distribution, are not valid for extremal QR. By using extreme value theory, which specifically accounts for the extreme nature of the tail data, we are able to provide inference methods that are valid for extremal QR.

Before describing the contributions of this paper in more detail, we first motivate the use of extremal quantile regression in specific economic applications. Extremal quantile regression provides a useful description of important features of the data in these applications, generating both reduced-form facts as well as inputs into estimation of structural models. In what follows,  $Q_Y(\tau|X)$  denotes the conditional  $\tau$ -quantile of  $Y$  given  $X$ ; extremal conditional quantile refers to the conditional quantile function  $Q_Y(\tau|X)$  with the quantile index  $\tau = \epsilon$  or  $1 - \epsilon$ , where  $\epsilon$  is close to zero; and extremal quantile regression refers to the quantile regression estimator of an extremal conditional quantile.

A principal area of economic applications of extremal quantile regression is risk management. One example in this area is conditional value-at-risk analysis from financial economics (Chernozhukov and Umansev 2001, Engle and Manganelli 2004). Here, we are interested in the extremal quantile  $Q_Y(\epsilon|X)$  of a return  $Y$  to a bank's portfolio, conditional on various predictive variables  $X$ , such as the return to the market portfolio and the returns to portfolios of other related banks and mortgage providers. Unlike unconditional extremal quantiles, conditional extremal quantiles are useful for stress testing and analyzing the impact of adverse systemic events on the bank's performance. For example, we can analyze the impact of a large drop in the value of the market portfolio or of an associated company on the performance of the bank's portfolio. The results of this analysis are useful for determining the level of capital that the bank needs to hold to prevent bankruptcy in unfavorable states of the world. Another example comes from health economics, where we are interested in the analysis of socio-economic determinants  $X$  of extreme quantiles of a child's birthweight  $Y$  or other health outcomes. In this example, very low birthweights are connected with substantial health problems for the child, and thus extremal quantile regression is useful to identifying which factors can improve these negative health outcomes. We shall return to these examples later in the empirical part of the paper.

Another primary area of economic applications of extremal quantile regression deals with describing approximate or probabilistic boundaries of economic outcomes conditional on pertinent factors. A first example in this area comes from efficiency analysis in the economics of regulation, where we are interested in the probabilistic production frontier  $Q_Y(1 - \epsilon|X)$ . This frontier describes the level of production  $Y$  attained by the most productive  $(1 - \epsilon) \times 100$  percent of firms, conditional on input factors  $X$  (Timmer 1971). A second example comes from the analysis of job search in labor economics, where we are interested in the approximate reservation wage  $Q_Y(\epsilon|X)$ . This function describes the wage level, below which the worker accepts a job only with a small probability  $\epsilon$ , conditional on worker characteristics and other factors  $X$  (Flinn and Heckman 1982). A third example deals with estimating  $(S, s)$  rules in industrial organization and macroeconomics (Caballero and Engel 1999). Recall that the  $(S, s)$  rule is an optimal policy for capital adjustment, in which a firm allows its capital stock to gradually depreciate to a lower barrier, and once the barrier is reached, the firm adjusts its capital stock sharply to an upper barrier. Therefore, in a given cross-section of firms, the extremal conditional quantile functions  $Q_Y(\epsilon|X)$  and  $Q_Y(1 - \epsilon|X)$  characterize the approximate adjustment barriers for observed capital stock  $Y$ , conditional on a set of observed factors  $X$ .<sup>1,2</sup>

The two areas of applications described above are either non-structural or semi-structural. A third principal area of economic applications of extremal quantile regressions is structural estimation of economic models. For instance, in procurement auction models, the key information about structural parameters is contained in the extreme or near-extreme conditional quantiles of bids given bidder and auction characteristics (see e.g. Chernozhukov and Hong (2004) and Hirano and Porter (2003)). We then can estimate and test a structural model based on its ability to accurately reproduce a collection of extremal conditional quantiles observed in the data. This indirect inference approach is called the method-of-quantiles (Koenker 2005). We refer the reader to Donald and Paarsch (2002) for a detailed example of this approach in the context of using  $k$ -sample extreme quantiles.<sup>3</sup>

---

<sup>1</sup>Caballero and Engel (1999) study approximate adjustment barriers using distribution models; obviously quantile models can also be used.

<sup>2</sup>In the previous examples, we can set  $\epsilon$  to 0 to recover the exact, non-probabilistic, boundaries in the case with no unobserved heterogeneity and no (even small) outliers in the data. Our inference methods cover this exact extreme case, but we recommend avoiding it because it requires very stringent assumptions.

<sup>3</sup>The method-of-quantiles allows us to estimate structural models both with and without parametric unobserved heterogeneity. Moreover, the use of near-extreme quantiles instead of exact-extreme quantiles makes the method more robust to a small fraction of outliers or neglected unobserved heterogeneity.

We now describe the contributions of this paper more specifically. This paper develops feasible and practical inferential methods based on extreme value (EV) theory for QR, namely, on the limit law theory for QR developed in Chernozhukov (2005) for cases where the quantile index  $\tau \in (0, 1)$  is either low, close to zero, or high, close to 1. Without loss of generality we assume the former. By close to 0, we mean that the order of the  $\tau$ -quantile,  $\tau T$ , defined as the product of quantile index  $\tau$  with the sample size  $T$ , obeys  $\tau T \rightarrow k < \infty$  as  $T \rightarrow \infty$ . Under this condition, the conventional normal laws, which are based on the assumption that  $\tau T$  diverges to infinity, fail to hold, and different EV laws apply instead. These laws approximate the exact finite sample law of extremal QR better than the conventional normal laws. In particular, we find that when the dimension-adjusted order of the  $\tau$ -quantile,  $\tau T/d$ , defined as the ratio of the order of the  $\tau$ -quantile to the number of regressors  $d$ , is not large, less than about 20 or 40, the EV laws may be preferable to the normal law, whereas the normal laws may become preferable otherwise. We suggest this simple rule of thumb for choosing between the EV laws and normal laws, and refer the reader to Section 5 for more refined suggestions and recommendations.

Figure 1 illustrates the difference between the EV and normal approximations to the finite sample distribution of the extremal QR estimators. We plot the quantiles of these approximations against the quantiles of the exact finite sample distribution of the QR estimator. We consider different dimension-adjusted orders in a simple model with only one regressor,  $d = 1$ , and  $T = 200$ . If either the EV law or the normal law were to coincide with the true law, then their quantiles would fall exactly on the 45 degree line shown by the solid line. We see from the plot that when the dimension-adjusted order  $\tau T/d$  is 20 or 40, the quantiles of the EV law are indeed very close to the 45 degree line, and in fact are much closer to this line than the quantiles of the normal law. Only for the case when the effective order  $\tau T/d$  becomes 60, do the quantiles of the EV law and normal laws become comparably close to the 45 degree line.

A major problem with implementing the EV approach, at least in its pure form, is its infeasibility for inference purposes. Indeed, EV approximations rely on canonical normalizing constants to achieve non-degenerate asymptotic laws. Consistent estimation of these constants is generally not possible, at least without making additional strong assumptions. This difficulty is also encountered in the classical non-regression case; see, for instance, Bertail, Haefke, Politis, and White (2004) for discussion. Furthermore, universal inference methods such as the bootstrap fail due to the nonstandard behavior of extremal QR statistics; see Bickel and Freedman (1981) for a proof in the classical non-regression case.

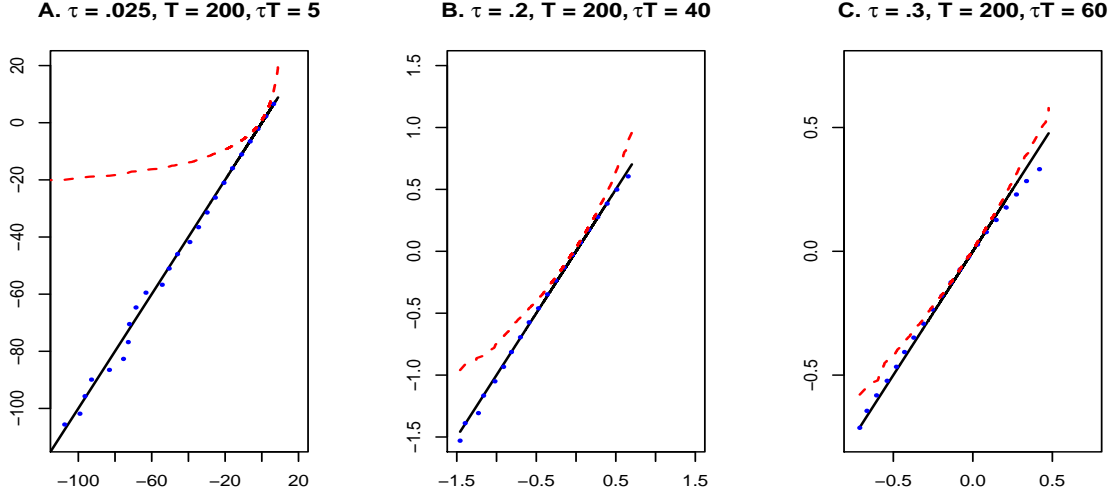


FIGURE 1. **Quantiles of the true law of QR vs. quantiles of EV and normal laws.** The figure is based on a simple design with  $Y = X + U$ , where  $U$  follows a Cauchy distribution and  $X = 1$ . The solid line “—” shows the actual quantiles of the true distribution of QR with quantile index  $\tau \in \{.025, .2, .3\}$ . The dashed line “- - -” shows the quantiles of the conventional normal law for QR, and the dotted line “.....” shows the quantiles of EV law for QR. The figure is based on 10,000 Monte Carlo replications and plots quantiles over the 99% range.

Conventional subsampling methods with and without replacement are also inconsistent because the QR statistic diverges in the unbounded support case. Moreover, they require consistent estimation of normalizing constants, which is not feasible in general.

In this paper we develop two types of inference approaches that overcome all of the difficulties mentioned above: a resampling approach and an analytical approach. We favor the first approach due to its ease of implementation in practice. At the heart of both approaches is the use of self-normalized QR (SN-QR) statistics that employ random normalization factors, instead of generally infeasible normalization by canonical constants. The use of SN-QR statistics allows us to derive feasible limit distributions, which underlie either of our inference approaches. Moreover, our resampling approach is a suitably modified subsampling method applied to SN-QR statistics. This approach entirely avoids estimating not only the canonical normalizing constants, but also all other nuisance tail parameters, which in practice may be difficult to estimate reliably. Our construction fruitfully exploits the special

relationship between the rates of convergence/divergence of extremal and intermediate QR statistics, which allows for a valid estimation of the centering constants in subsamples. For completeness we also provide inferential methods for canonically-normalized QR (CN-QR) statistics, but we also show that their feasibility requires much stronger assumptions.

The remainder of the paper is organized as follows. Section 2 describes the model and regularity conditions, and gives an intuitive overview of the main results. Section 3 establishes the results that underlie the inferential procedures. Section 4 describes methods for estimating critical values. Section 5 compares inference methods based on EV and normal approximations through a Monte Carlo experiment. Section 6 presents empirical examples, and the Appendix collects proofs and all other figures.

## 2. THE SET UP AND OVERVIEW OF RESULTS

**2.1. Some Basics.** Let a real random variable  $Y$  have a continuous distribution function  $F_Y(y) = \Pr[Y \leq y]$ . A  $\tau$ -quantile of  $Y$  is  $Q_Y(\tau) = \inf\{y : F_Y(y) > \tau\}$  for some  $\tau \in (0, 1)$ . Let  $X$  be a vector of covariates related to  $Y$ , and  $F_Y(y|x) = \Pr[Y \leq y|X = x]$  denote the conditional distribution function of  $Y$  given  $X = x$ . The conditional  $\tau$ -quantile of  $Y$  given  $X = x$  is  $Q_Y(\tau|x) = \inf\{y : F_Y(y|x) > \tau\}$  for some  $\tau \in (0, 1)$ . We refer to  $Q_Y(\tau|x)$ , viewed as a function of  $x$ , as the  $\tau$ -quantile regression function. This function measures the effect of covariates on outcomes, both at the center and at the upper and lower tails of the outcome distribution. A conditional  $\tau$ -quantile is *extremal* whenever the probability index  $\tau$  is either low or high in a sense that we will make more precise below. Without loss of generality, we focus the discussion on low quantiles.

Consider the classical linear functional form for the conditional quantile function of  $Y$  given  $X = x$ :

$$Q_Y(\tau|x) = x'\beta(\tau), \quad \text{for all } \tau \in \mathcal{I} = (0, \eta], \text{ for some } \eta \in (0, 1], \quad (2.1)$$

and for every  $x \in \mathbf{X}$ , the support of  $X$ . This linear functional form is flexible in the sense that it has good approximation properties. Indeed, given an original regressor  $X^*$ , the final set of regressors  $X$  can be formed as a vector of approximating functions. For example,  $X$  may include power functions, splines, and other transformations of  $X^*$ .

Given a sample of  $T$  observations  $\{Y_t, X_t, t = 1, \dots, T\}$ , the  $\tau$ -quantile QR estimator  $\hat{\beta}(\tau)$  solves:

$$\hat{\beta}(\tau) \in \arg \min_{\beta \in \mathbb{R}^d} \sum_{t=1}^T \rho_{\tau}(Y_t - X_t'\beta), \quad (2.2)$$

where  $\rho_\tau(u) = (\tau - 1(u < 0))u$  is the asymmetric absolute deviation function. The median  $\tau = 1/2$  case of (2.2) was introduced by Laplace (1818) and the general quantile formulation (2.2) by Koenker and Bassett (1978).

QR coefficients  $\hat{\beta}(\tau)$  can be seen as order statistics in the regression setting. Accordingly, we will refer to  $\tau T$  as the order of the  $\tau$ -quantile. A sequence of quantile index-sample size pairs  $\{\tau_T, T\}_{T=1}^\infty$  is said to be an *extreme order* sequence if  $\tau_T \searrow 0$  and  $\tau_T T \rightarrow k \in (0, \infty)$  as  $T \rightarrow \infty$ ; an *intermediate order* sequence if  $\tau_T \searrow 0$  and  $\tau_T T \rightarrow \infty$  as  $T \rightarrow \infty$ ; and a *central order* sequence if  $\tau_T$  is fixed as  $T \rightarrow \infty$ . Each type of sequence leads to different asymptotic approximations to the finite-sample distribution of the QR estimator. The extreme order sequence leads to an extreme value (EV) law in large samples, whereas the intermediate and central sequences lead to normal laws. As we saw in Figure 1, the EV law provides a better approximation to the finite sample law of the QR estimator than the normal law.

**2.2. Pareto-type or Regularly Varying Tails.** In order to develop inference theory for extremal QR, we assume the tails of the conditional distribution of the outcome variable have Pareto-type behavior, as we formally state in the next subsection. In this subsection, we recall and discuss the concept of Pareto-type tails. The (lower) tail of a distribution function has Pareto-type behavior if it decays approximately as a power function, or more formally, a regularly varying function. The tails of the said form are prevalent in economic data, as discovered by V. Pareto in 1895.<sup>4</sup> Pareto-type tails encompass or approximate a rich variety of tail behavior, including that of thick-tailed and thin-tailed distributions, having either bounded or unbounded support.

More formally, consider a random variable  $Y$  and define a random variable  $U$  as  $U \equiv Y$ , if the lower end-point of the support of  $Y$  is  $-\infty$ , or  $U \equiv Y - Q_Y(0)$ , if the lower end-point of the support of  $Y$  is  $Q_Y(0) > -\infty$ . The quantile function of  $U$ , denoted by  $Q_U$ , then has lower end-point  $Q_U(0) = -\infty$  or  $Q_U(0) = 0$ . The assumption that the quantile function  $Q_U$  and its distribution function  $F_U$  exhibit Pareto-type behavior in the tails can be formally stated as the following two equivalent conditions:<sup>5</sup>

$$Q_U(\tau) \sim L(\tau) \cdot \tau^{-\xi} \quad \text{as } \tau \searrow 0, \quad (2.3)$$

$$F_U(u) \sim \bar{L}(u) \cdot u^{-1/\xi} \quad \text{as } u \searrow Q_U(0), \quad (2.4)$$

<sup>4</sup> Pareto called the tails of this form “A Distribution Curve for Wealth and Incomes.” Further empirical substantiation has been given by Sen (1973), Zipf (1949), Mandelbrot (1963), and Fama (1965), among others. The mathematical theory of regular variation in connection to extreme value theory has been developed by Karamata, Gnedenko, and de Haan.

<sup>5</sup> The notation  $a \sim b$  means that  $a/b \rightarrow 1$  as appropriate limits are taken.

for some real number  $\xi \neq 0$ , where  $L(\tau)$  is a nonparametric slowly-varying function at 0, and  $\bar{L}(u)$  is a nonparametric slowly-varying function at  $Q_U(0)$ .<sup>6</sup> The leading examples of slowly-varying functions are the constant function and the logarithmic function. The number  $\xi$  defined in (2.3) and (2.3) is called the EV index.

The absolute value of  $\xi$  measures the heavy-tailedness of the distribution. A distribution  $F_Y$  with Pareto-type tails necessarily has a finite lower support point if  $\xi < 0$  and a infinite lower support point if  $\xi > 0$ . Distributions with  $\xi > 0$  include stable, Pareto, Student's  $t$ , and many other distributions. For example, the  $t$  distribution with  $\nu$  degrees of freedom has  $\xi = 1/\nu$  and exhibits a wide range of tail behavior. In particular, setting  $\nu = 1$  yields the Cauchy distribution which has heavy tails with  $\xi = 1$ , while setting  $\nu = 30$  gives a distribution which has light tails with  $\xi = 1/30$ , and which is very close to the normal distribution. On the other hand, distributions with  $\xi < 0$  include the uniform, exponential, Weibull, and many other distributions.

It should be mentioned that the case of  $\xi = 0$  corresponds to the class of rapidly varying distribution functions. These distribution functions have exponentially light tails, with the normal and exponential distributions being the chief examples. To simplify the exposition, we do not discuss this case explicitly. However, since the limit distributions of the main statistics are continuous in  $\xi$ , including at  $\xi = 0$ , inference theory for the case of  $\xi = 0$  is also included by taking  $\xi \rightarrow 0$ .

**2.3. The Extremal Conditional Quantile Model and Sampling Conditions.** With these notions in mind, our main assumption is that the response variable  $Y$ , transformed by some auxiliary regression line,  $X'\beta_e$ , has Pareto-type tails with EV index  $\xi$ .

**C1.** *The conditional quantile function of  $Y$  given  $X = x$  satisfies equation (2.1) a.s. Moreover, there exists an auxiliary extremal regression parameter  $\beta_e \in \mathbb{R}^d$ , such that the disturbance  $V \equiv Y - X'\beta_e$  has end-point  $s = 0$  or  $s = -\infty$  a.s., and its conditional quantile function  $Q_V(\tau|x)$  satisfies the following tail-equivalence relationship:*

$$Q_V(\tau|x) \sim x'\gamma \cdot Q_U(\tau), \text{ as } \tau \searrow 0, \text{ uniformly in } x \in \mathbf{X} \subseteq \mathbb{R}^d,$$

*for some quantile function  $Q_U(\tau)$  that exhibits Pareto-type tails with EV index  $\xi$  (i.e., it satisfies (2.3)), and some vector parameter  $\gamma$  such that  $E[X]'\gamma = 1$ .*

Since this assumption only affects the tails, it allows covariates to affect the extremal quantile and the central quantiles very differently. Moreover, the local effect of covariates

---

<sup>6</sup>A function  $u \mapsto L(u)$  is said to be slowly-varying at  $s$  if  $\lim_{l \searrow s} [L(l)/L(ml)] = 1$  for any  $m > 0$ .



in the tail is approximately given by  $\beta(\tau) \approx \beta_e + \gamma Q_U(\tau)$ , which allows for a differential impact of covariates across various extremal quantiles.

**C2.** *The conditional quantile density function  $\partial Q_V(\tau|x)/\partial\tau$  exists and satisfies the tail equivalence relationship  $\partial Q_V(\tau|x)/\partial\tau \sim x'\gamma \cdot \partial Q_U(\tau)/\partial\tau$  as  $\tau \searrow 0$ , uniformly in  $x \in \mathbf{X}$ , where  $\partial Q_U(\tau)/\partial\tau$  exhibits Pareto-type tails as  $\tau \searrow 0$  with EV index  $\xi + 1$ .*

Assumption C2 strengthens C1 by imposing the existence and Pareto-type behavior of the conditional quantile density function. We impose C2 to facilitate the derivation of the main inferential results.

The following sampling conditions will be imposed.

**C3.** *The regressor vector  $X = (1, Z')'$  is such that it has a compact support  $\mathbf{X}$ , the matrix  $E[XX']$  is positive definite, and its distribution function  $F_X$  satisfies a non-lattice condition stated in the mathematical appendix (this condition is satisfied, for instance, when  $Z$  is absolutely continuous).*

Compactness is needed to ensure the continuity and robustness of the mapping from extreme events in  $Y$  to the extremal QR statistics. Even if  $X$  is not compact, we can select the data for which  $X$  belongs to a compact region. The non-degeneracy condition of  $E[XX']$  is standard and guarantees invertibility. The non-lattice condition is required for the existence of the finite-sample density of QR coefficients. It is needed even asymptotically because the asymptotic distribution theory of extremal QR closely resembles the finite-sample theory for QR, which is not a surprise given the rare nature of events that have a probability of order  $1/T$ .

We assume the data are either i.i.d. or weakly dependent.

**C4.** *The sequence  $\{W_t\}$  with  $W_t = (V_t, X_t)$  and  $V_t$  defined in C1, forms a stationary, strongly mixing process with a geometric mixing rate, that is, for some  $C > 0$*

$$\sup_t \sup_{A \in \mathcal{A}_t, B \in \mathcal{B}_{t+m}} |P(A \cap B) - P(A)P(B)| \exp(Cm) \rightarrow 0 \text{ as } m \rightarrow \infty,$$

where  $\mathcal{A}_t = \sigma(W_t, W_{t-1}, \dots)$  and  $\mathcal{B}_t = \sigma(W_t, W_{t+1}, \dots)$ . Moreover, the sequence satisfies a condition that curbs clustering of extreme events in the following sense:  $P(V_t \leq K, V_{t+j} \leq K | \mathcal{A}_t) \leq CP(V_t \leq K | \mathcal{A}_t)^2$  for all  $K \in [s, \bar{K}]$ , uniformly for all  $j \geq 1$  and uniformly for all  $t \geq 1$ ; here  $C > 0$  and  $\bar{K} > s$  are some constants.

A special case of this condition is when the sequence of variables  $\{(V_t, X_t), t \geq 1\}$ , or equivalently  $\{(Y_t, X_t), t \geq 1\}$ , is independent and identically distributed. The assumption

of mixing for  $\{(V_t, X_t), t \geq 1\}$  is standard in econometric analysis (White 1990), and it is equivalent to the assumption of mixing of  $\{(Y_t, X_t), t \geq 1\}$ . The non-clustering condition is of the Meyer (1973)-type and states that the probability of two extreme events co-occurring at nearby dates is much lower than the probability of just one extreme event. For example, it assumes that a large market crash is not likely to be immediately followed by another large crash. This assumption leads to limit distributions of QRs as if independent sampling had taken place. The plausibility of the non-clustering assumption is an empirical matter. We conjecture that our primary inference method based on subsampling is valid more generally, under conditions that preserve the rates of convergence of QR statistics and ensure existence of their asymptotic distributions. Finally we note that the assumptions made here could be relaxed in certain directions for some of the results stated below, but we decided to state a single set of sufficient assumptions for all the results.

**2.4. Overview and Discussion of Inferential Results.** We begin by briefly revisiting the classical non-regression case to describe some intuition and the key obstacles to performing feasible inference in our more general regression case. Then we will describe our main inferential results for the regression case. It is worth noting that our main inferential methods, based on self-normalized statistics, are new and of independent interest even in the classical non-regression case.

Recall the following classical result on the limit distribution of the extremal sample quantiles  $\widehat{Q}_Y(\tau)$  (Gnedenko 1943): for any integer  $k \geq 1$  and  $\tau = k/T$ , as  $T \rightarrow \infty$ ,

$$\widehat{Z}_T(k) = A_T(\widehat{Q}_Y(\tau) - Q_Y(\tau)) \rightarrow_d \widehat{Z}_\infty(k) = \Gamma_k^{-\xi} - k^{-\xi}, \quad (2.5)$$

where

$$A_T = 1/Q_U(1/T), \quad \Gamma_k = \mathcal{E}_1 + \dots + \mathcal{E}_k, \quad (2.6)$$

and  $(\mathcal{E}_1, \mathcal{E}_2, \dots)$  is an independent and identically distributed sequence of standard exponential variables. We refer to  $\widehat{Z}_T(k)$  as the canonically normalized (CN) statistic because it depends on the scaling constant  $A_T$ . The variables  $\Gamma_k$ , entering the definition of the EV distribution, are *gamma* random variables. The limit distribution of the  $k$ -th order statistic is therefore a transformation of a gamma variable. The EV distribution is not symmetric and may have significant (median) bias; it has finite moments if  $\xi < 0$  and has finite moments of up to order  $1/\xi$  if  $\xi > 0$ . The presence of median bias motivates the use of median-bias correction techniques, which we discuss in the regression case below.

Although very powerful, this classical result is not feasible for purposes of inference on  $Q_Y(\tau)$ , since the scaling constant  $A_T$  is generally not possible to estimate consistently (Bertail, Haefke, Politis, and White 2004). One way to deal with this problem is to add strong parametric assumptions on the non-parametric, slowly-varying function  $L(\cdot)$  in equation (2.3) in order to estimate  $A_T$  consistently. For instance, suppose that  $Q_U(\tau) \sim L\tau^{-\xi}$ . Then one can estimate  $\xi$  by the classical Hill or Pickands estimators, and  $L$  by  $\hat{L} = (\hat{Q}_Y(2\tau) - \hat{Q}_Y(\tau))/(2^{-\hat{\xi}} - 1)\tau^{-\hat{\xi}}$ . We develop the necessary theoretical results for the regression analog of this approach, although we will not recommend it as our preferred method.

Our preferred and main proposal to deal with the aforementioned infeasibility problem is to consider the asymptotics of the self-normalized (SN) sample quantiles

$$Z_T(k) = \mathcal{A}_T(\hat{Q}_Y(\tau) - Q_Y(\tau)) \rightarrow_d Z_\infty(k) = \frac{\sqrt{k}(\Gamma_k^{-\xi} - k^{-\xi})}{\Gamma_{mk}^{-\xi} - \Gamma_k^{-\xi}}, \quad (2.7)$$

where for  $m > 1$  such that  $mk$  is an integer,

$$\mathcal{A}_T = \frac{\sqrt{\tau T}}{\hat{Q}_Y(m\tau) - \hat{Q}_Y(\tau)}. \quad (2.8)$$

Here, the scaling factor  $\mathcal{A}_T$  is completely a function of data and therefore *feasible*. Moreover, we completely avoid the need for consistent estimation of  $A_T$ . This is convenient because we are not interested in this normalization constant per se. The limit distribution in (2.7) only depends on the EV index  $\xi$ , and its quantiles can be easily obtained by simulation. In the regression setting, where the limit law is a bit more complicated, we develop a form of subsampling to perform both practical and feasible inference.

Let us now turn to the regression case. Here, we can also consider a canonically-normalized QR statistic (CN-QR):

$$\hat{Z}_T(k) := A_T(\hat{\beta}(\tau) - \beta(\tau)) \text{ for } A_T := 1/Q_U(1/T); \quad (2.9)$$

and a self-normalized QR (SN-QR) statistic:

$$Z_T(k) := \mathcal{A}_T(\hat{\beta}(\tau) - \beta(\tau)) \text{ for } \mathcal{A}_T := \frac{\sqrt{\tau T}}{\bar{X}'_T(\hat{\beta}(m\tau) - \hat{\beta}(\tau))}, \quad (2.10)$$

where  $\bar{X}_T = \sum_{t=1}^T X_t/T$  and  $m$  is a real number such that  $\tau T(m-1) > d$ . The first statistic uses an infeasible canonical normalization  $A_T$ , whereas the second statistic uses a feasible random normalization. First, we show that

$$\hat{Z}_T(k) \rightarrow_d \hat{Z}_\infty(k) \quad (2.11)$$

where for  $\chi = 1$  if  $\xi < 0$  and  $\chi = -1$  if  $\xi > 0$ ,

$$\widehat{Z}_\infty(k) := \chi \cdot \arg \min_{z \in \mathbb{R}^d} \left[ -kE[X]'(z + k^{-\xi}\gamma) + \sum_{t=1}^{\infty} \{\mathcal{X}'_t(z + k^{-\xi}\gamma) - \chi \cdot \Gamma_t^{-\xi} \cdot \mathcal{X}'_t\gamma\}_+ \right] \quad (2.12)$$

where  $\{\Gamma_1, \Gamma_2, \dots\} := \{\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, \dots\}$ ;  $\{\mathcal{E}_1, \mathcal{E}_2, \dots\}$  is an iid sequence of exponential variables that is independent of  $\{\mathcal{X}_1, \mathcal{X}_2, \dots\}$ , an iid sequence with distribution  $F_X$ ; and  $\{y\}_+ := \max(0, y)$ . Furthermore, we show that

$$Z_T(k) \rightarrow_d Z_\infty(k) := \frac{\sqrt{k}\widehat{Z}_\infty(k)}{E[X]'(\widehat{Z}_\infty(mk) - \widehat{Z}_\infty(k)) + \chi \cdot (m^{-\xi} - 1)k^{-\xi}}. \quad (2.13)$$

The limit laws here are more complicated than in the non-regression case, but they share some common features. Indeed, the limit laws depend on the variables  $\Gamma_i$  in a crucial way, and are not necessarily centered at zero and can have significant first order median biases. Motivated by the presence of the first order bias, we develop bias corrections for the QR statistics in the next section. Moreover, just as in the non-regression case, the limit distribution of the CN-QR statistic in (2.12) is generally infeasible for inference purposes. We need to know or estimate the scaling constant  $A_T$ , which is the reciprocal of the extremal quantile of the variable  $U$  defined in C1. That is, we require an estimator  $\widehat{A}_T$  such that  $\widehat{A}_T/A_T \rightarrow_p 1$ , which is not feasible unless the tail of  $U$  satisfies additional strong parametric restrictions. We provide additional restrictions below that facilitate estimation of  $A_T$  and hence inference based on CN-QR, although this is not our preferred inferential method.

Our main and preferred proposal for inference is based on the SN-QR statistic, which does not depend on  $A_T$ . We estimate the distribution of this statistic using either a variation of subsampling or an analytical method. A key ingredient here is the feasible normalizing variable  $\mathcal{A}_T$ , which is randomly proportional to the canonical normalization  $A_T$ , in the sense that  $\mathcal{A}_T/A_T$  is a random variable in the limit.<sup>7</sup> An advantage of the subsampling method over the analytical methods is that it does not require estimation of the nuisance parameters  $\xi$  and  $\gamma$ . Our subsampling approach is different from conventional subsampling in the use of recentering terms and random normalization. Conventional subsampling that uses recentering by the full sample estimate  $\widehat{\beta}(\tau)$  is not consistent when that estimate is diverging; and here we indeed have  $A_T \rightarrow 0$  when  $\xi > 0$ . Instead, we recenter by intermediate order

---

<sup>7</sup>The idea of feasible random normalization has been used in other contexts (e.g. t-statistics). In extreme value theory, Dekkers and de Haan (1989) applied a similar random normalization idea to extrapolated quantile estimators of intermediate order in the non-regression setting, precisely to produce limit distributions that can be easily used for inference. In time series, Kiefer, Vogelsang, and Bunzel (2000) have used feasible inconsistent estimates of the variance of asymptotically normal estimators.

QR estimates in subsamples, which will diverge at a slow enough speed to estimate the limit distribution of SN-QR consistently. Thus, our subsampling approach explores the special relationship between the rates of convergence/divergence of extremal and intermediate QR statistics and should be of independent interest even in a non-regression setting.

This paper contributes to the existing literature by introducing general feasible inference methods for extremal quantile regression. Our inferential methods rely in part on the limit results in Chernozhukov (2005), who derived EV limit laws for CN-QR under the extreme order condition  $\tau T \rightarrow k > 0$ . This theory, however, did not lead directly to any feasible, practical inference procedure. Feigin and Resnick (1994), Chernozhukov (1998), Portnoy and Jurečková (1999), and Knight (2001) provide related limit results for canonically normalized linear programming estimators where  $\tau T \searrow 0$ , all in different contexts and at various levels of generality. These limit results likewise did not provide feasible inference theory. The linear programming estimator is well suited to the problem of estimating finite deterministic boundaries of data, as in image processing and other technometric applications. In contrast, the current approach of taking  $\tau T \rightarrow k > 0$  is more suited to econometric applications, where interest focuses on the “usual” quantiles located near the minimum or maximum and where the boundaries may be unlimited. However, some of our theoretical developments are motivated by and build upon this previous literature. Some of our proofs rely on the elegant epi-convergence framework of Geyer (1996) and Knight (1999).

### 3. INFERENCE AND MEDIAN-UNBIASED ESTIMATION BASED ON EXTREME VALUE LAWS

This section establishes the main results that underlie our inferential procedures.

**3.1. Extreme Value Laws for CN-QR and SN-QR Statistics.** Here we verify that that the CN-QR statistic  $\hat{Z}_T(k)$  and SN-QR statistic  $Z_T(k)$  converge to the limit variables  $\hat{Z}_\infty(k)$  and  $Z_\infty(k)$ , under the condition that  $\tau T \rightarrow k > 0$  as  $T \rightarrow \infty$ .

**Theorem 1** (Limit Laws for Extremal SN-QR and CN-QR). *Suppose conditions C1, C3 and C4 hold. Then as  $\tau T \rightarrow k > 0$  and  $T \rightarrow \infty$ , (1) the SN-QR statistic of order  $k$  obeys*

$$Z_T(k) \rightarrow_d Z_\infty(k),$$

*for any  $m$  such that  $k(m-1) > d$ , and (2) the CN-QR statistic of order  $k$  obeys*

$$\hat{Z}_T(k) \rightarrow_d \hat{Z}_\infty(k).$$

**Comment 3.1.** The condition that  $k(m-1) > d$  in the definition of SN-QR ensures that  $\beta(m\tau) \neq \beta(\tau)$  and therefore the normalization by  $\mathcal{A}_T$  is well defined. This is a consequence of Theorem 3.2 in Bassett and Koenker (1982) and existence of the conditional density of

$Y$  imposed in assumption C2. Result 1 on SN-QR statistics is the main new result that we will exploit for inference. Result 2 on CN-QR statistics is needed primarily for auxiliary purposes. Chernozhukov (2005) presents some extensions of result 2.

**Comment 3.2.** When  $Q_Y(0|x) > -\infty$ , by C1  $Q_Y(0|x)$  is equal to  $x'\beta_e$  and is the conditional lower boundary of  $Y$ . The proof of Theorem 1 shows that

$$A_T(\hat{\beta}(\tau) - \beta_e) \rightarrow_d \tilde{Z}_\infty(k) := \hat{Z}_\infty(k) - k^{-\xi} \text{ and } \mathcal{A}_T(\hat{\beta}(\tau) - \beta_e) \rightarrow_d \tilde{Z}_\infty(k) / (\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k)).$$

We can use these results and analytical and subsampling methods presented below to perform median unbiased estimation and inference on the boundary parameter  $\beta_e$ .<sup>8</sup>

**3.2. Generic Inference and Median-Unbiased Estimation.** We outline two procedures for conducting inference and constructing asymptotically median unbiased estimates of linear functions  $\psi'\beta(\tau)$  of the coefficient vector  $\beta(\tau)$ , for some non-zero vector  $\psi$ .

**1. Median-Unbiased Estimation and Inference Using SN-QR.** By Theorem 1,  $\psi' A_T(\hat{\beta}(\tau) - \beta(\tau)) \rightarrow_d \psi' Z_\infty(k)$ . Let  $c_\alpha$  denote the  $\alpha$ -quantile of  $\psi' Z_\infty(k)$  for  $0 < \alpha \leq .5$ . Given  $\hat{c}_\alpha$ , a consistent estimate of  $c_\alpha$ , we can construct an asymptotically median-unbiased estimator and a  $(1 - \alpha)\%$ -confidence interval for  $\psi'\beta(\tau)$  as

$$\psi'\hat{\beta}(\tau) - \hat{c}_{1/2}/\mathcal{A}_T \text{ and } [\psi'\hat{\beta}(\tau) - \hat{c}_{1-\alpha/2}/\mathcal{A}_T, \psi'\hat{\beta}(\tau) - \hat{c}_{\alpha/2}/\mathcal{A}_T],$$

respectively. The bias-correction term and the limits of the confidence interval depend on the random scaling  $\mathcal{A}_T$ . We provide consistent estimates of  $c_\alpha$  in the next section.

**Theorem 2** (Inference and median-unbiased estimation using SN-QR). *Under the conditions of Theorem 1, suppose we have  $\hat{c}_\alpha$  such that  $\hat{c}_\alpha \rightarrow_p c_\alpha$ . Then,*

$$\lim_{T \rightarrow \infty} P\{\psi'\hat{\beta}(\tau) - \hat{c}_{1/2}/\mathcal{A}_T \leq \psi'\beta(\tau)\} = 1/2$$

and

$$\lim_{T \rightarrow \infty} P\{\psi'\hat{\beta}(\tau) - \hat{c}_{1-\alpha/2}/\mathcal{A}_T \leq \psi'\beta(\tau) \leq \psi'\hat{\beta}(\tau) - \hat{c}_{\alpha/2}/\mathcal{A}_T\} = 1 - \alpha.$$

**2. Median Unbiased Estimation and Inference Using CN-QR.** By Theorem 1,  $\psi' A_T(\hat{\beta}(\tau) - \beta(\tau)) \rightarrow_d \psi' \hat{Z}_\infty(k)$ . Let  $c'_\alpha$  denote the  $\alpha$ -quantile of  $\psi' \hat{Z}_\infty(k)$  for  $0 < \alpha \leq .5$ . Given  $\hat{A}_T$ , a consistent estimate of  $A_T$ , and  $\hat{c}'_\alpha$ , a consistent estimate of  $c'_\alpha$ , we can construct

---

<sup>8</sup> To estimate the critical values, we can use either analytical or subsampling methods presented below, with the difference that in subsampling we need to recenter by the full sample estimate  $\hat{\beta}_e = \hat{\beta}(1/T)$ .

an asymptotically median-unbiased estimator and a  $(1 - \alpha)\%$ -confidence interval for  $\psi'\beta(\tau)$  as

$$\psi'\hat{\beta}(\tau) - \hat{c}'_{1/2}/\hat{A}_T \quad \text{and} \quad [\psi'\hat{\beta}(\tau) - \hat{c}'_{1-\alpha/2}/\hat{A}_T, \psi'\hat{\beta}(\tau) - \hat{c}'_{\alpha/2}/\hat{A}_T],$$

respectively.

As mentioned in Section 2, construction of consistent estimates of  $A_T$  requires additional strong restrictions on the underlying model as well as additional steps in estimation. For example, suppose the nonparametric slowly varying component  $L(\tau)$  of  $A_T$  is replaced by a constant  $L$ , i.e. suppose that as  $\tau \searrow 0$

$$1/Q_U(\tau) = L \cdot \tau^\xi \cdot (1 + \delta(\tau)) \quad \text{for some } L \in \mathbb{R}, \text{ where } \delta(\tau) \rightarrow 0. \quad (3.14)$$

We can estimate the constants  $L$  and  $\xi$  via Pickands-type procedures:

$$\hat{\xi} = \frac{-1}{\ln 2} \ln \frac{\bar{X}'_T(\hat{\beta}(4\tau_T) - \hat{\beta}(\tau_T))}{\bar{X}'_T(\hat{\beta}(2\tau_T) - \hat{\beta}(\tau_T))} \quad \text{and} \quad \hat{L} = \frac{\bar{X}'_T(\hat{\beta}(2\tau_T) - \hat{\beta}(\tau_T))}{(2^{-\hat{\xi}} - 1) \cdot \tau^{-\hat{\xi}}}, \quad (3.15)$$

where  $\tau_T$  is chosen to be of an intermediate order,  $\tau_T T \rightarrow \infty$  and  $\tau_T \rightarrow 0$ . Theorem 4 in Chernozhukov (2005) shows that under C1-C4, condition (3.14), and additional conditions on the sequence  $(\delta(\tau_T), \tau_T)$ ,<sup>9</sup>  $\hat{\xi} = \xi + o(1/\ln T)$  and  $\hat{L} \rightarrow_p L$ , which produces the required consistent estimate  $\hat{A}_T = \hat{L}(1/T)^{-\hat{\xi}}$  such that  $\hat{A}_T/A_T \rightarrow_p 1$ . These additional conditions on the tails of  $Y$  and on the sequence  $(\delta(\tau_T), \tau_T)$  highlight the drawbacks of this inference approach relative to the previous one.

We provide consistent estimates of  $c'_\alpha$  in the next section.

**Theorem 3** (Inference and Median-Unbiased Estimation using CN-QR). *Assume the conditions of Theorem 1 hold. Suppose that we have  $\hat{A}_T$  such that  $\hat{A}_T/A_T \rightarrow_p 1$  and  $\hat{c}'_\alpha$  such that  $\hat{c}_\alpha \rightarrow_p c'_\alpha$ . Then,*

$$\lim_{T \rightarrow \infty} P\{\psi'\hat{\beta}(\tau) - \hat{c}'_{1/2}/\hat{A}_T \leq \psi'\beta(\tau)\} = 1/2$$

and

$$\lim_{T \rightarrow \infty} P\{\psi'\hat{\beta}(\tau) - \hat{c}'_{1-\alpha/2}/\hat{A}_T \leq \psi'\beta(\tau) \leq \psi'\hat{\beta}(\tau) - \hat{c}'_{\alpha/2}/\hat{A}_T\} = 1 - \alpha.$$

---

<sup>9</sup> The rate convergence of  $\hat{\xi}$  is  $\max[\frac{1}{\sqrt{\tau_T T}}, \ln \delta(\tau_T)]$ , which gives the following condition on the sequence  $(\delta(\tau_T), \tau_T)$  :  $\max[\frac{1}{\sqrt{\tau_T T}}, \ln \delta(\tau_T)] = o(1/\ln T)$ .

## 4. ESTIMATION OF CRITICAL VALUES

**4.1. Subsampling-Based Estimation of Critical Values.** Our resampling method for inference uses subsamples to estimate the distribution of SN-QR, as in standard subsampling. However, in contrast to the subsampling, our method bypasses estimation of the unknown convergence rate  $A_T$  by using self-normalized statistics. Our method also employs a special recentering that allows us to avoid the inconsistency of standard subsampling due to diverging QR statistics when  $\xi > 0$ .

The method has the following steps. First, consider all subsets of the data  $\{W_t = (Y_t, X_t), t = 1, \dots, T\}$  of size  $b$ ; if  $\{W_t\}$  is a time series, consider  $B_T = T - b + 1$  subsets of size  $b$  of the form  $\{W_i, \dots, W_{i+b-1}\}$ . Then compute the analogs of the SN-QR statistic, denoted  $\widehat{V}_{i,b}$  and defined below in equation (4.17), for each  $i$ -th subsample for  $i = 1, \dots, B_T$ . Second, obtain  $\widehat{c}_\alpha$  as the sample  $\alpha$ -quantile of  $\{\widehat{V}_{i,b,T}, i = 1, \dots, B_T\}$ . In practice, a smaller number  $B_T$  of randomly chosen subsets can be used, provided that  $B_T \rightarrow \infty$  as  $T \rightarrow \infty$ . (See Section 2.5 in Politis, Romano, and Wolf (1999).) Politis, Romano, and Wolf (1999) and Bertail, Haefke, Politis, and White (2004) provide rules for the choice of subsample size  $b$ .

The SN-QR statistic for the full sample of size  $T$  is:

$$V_T := \mathcal{A}_T \psi'(\widehat{\beta}_T(\tau_T) - \beta(\tau_T)) \text{ for } \mathcal{A}_T = \frac{\sqrt{\tau_T T}}{\bar{X}'_T(\widehat{\beta}(m\tau_T) - \widehat{\beta}(\tau_T))}, \quad (4.16)$$

where we can set  $m = (d + p)/(\tau_T T) + 1 = (d + p)/k + 1 + o(1)$ , where  $p \geq 1$  is the spacing parameter, which we set to 5.<sup>10</sup> In this section we write  $\tau_T$  to emphasize the theoretical dependence of the quantile of interest  $\tau$  on the sample size. In each  $i$ -th subsample of size  $b$ , we compute the following analog of  $V_T$ :

$$\widehat{V}_{i,b,T} := \mathcal{A}_{i,b,T} \psi'(\widehat{\beta}_{i,b,T}(\tau_b) - \widehat{\beta}(\tau_b)) \text{ for } \mathcal{A}_{i,b,T} := \frac{\sqrt{\tau_b b}}{\bar{X}'_{i,b,T}(\widehat{\beta}_{i,b,T}(m\tau_b) - \widehat{\beta}_{i,b,T}(\tau_b))}, \quad (4.17)$$

where  $\widehat{\beta}(\tau)$  is the  $\tau$ -quantile regression coefficient computed using the full sample,  $\widehat{\beta}_{i,b,T}(\tau)$  is the  $\tau$ -quantile regression coefficient computed using the  $i$ -th subsample,  $\bar{X}_{i,b,T}$  is the sample mean of the regressors in the  $i$ th subsample, and  $\tau_b := (\tau_T T)/b$ .<sup>11</sup> The determination of  $\tau_b$  is a critical decision that sets apart the extremal order approximation from the central order approximation. In the latter case, one sets  $\tau_b = \tau_T$  in subsamples. In the extreme

<sup>10</sup>Variation of this parameter from  $p = 2$  to  $p = 20$  yielded similar results in our Monte-Carlo experiments.

<sup>11</sup>In practice, it is reasonable to use the following finite-sample adjustment to  $\tau_b$ :  $\tau_b = \min[(\tau_T T)/b, .2]$  if  $\tau_T < .2$ , and  $\tau_b = \tau_T$  if  $\tau_T \geq .2$ . The idea is that  $\tau_T$  is judged to be non-extremal if  $\tau_T > .2$ , and the subsampling procedure reverts to central order inference. The truncation of  $\tau_b$  by  $.2$  is a finite-sample adjustment that restricts the key statistics  $\widehat{V}_{i,b,T}$  to be extremal in subsamples. These finite-sample adjustments do not affect the asymptotic arguments.



order approximation, our choice of  $\tau_b$  gives the same extreme order of  $\tau_b b$  in the subsample as the order of  $\tau_T T$  in the full sample.

Under the additional parametric assumptions on the tail behavior stated earlier, we can estimate the quantiles of the limit distribution of CN-QR using the following procedure: First, create subsamples  $i = 1, \dots, B_T$  as before and compute in each subsample:  $\tilde{V}_{i,b,T} := \hat{A}_b \psi'(\hat{\beta}_{i,b,T}(\tau_b) - \hat{\beta}(\tau_b))$ , where  $\hat{A}_b$  is any consistent estimate of  $A_b$ . For example, under the parametric restrictions specified in (3.14), set  $\hat{A}_b = \hat{L}b^{-\hat{\xi}}$  for  $\hat{L}$  and  $\hat{\xi}$  specified in (3.15). Second, obtain  $\hat{c}'_\alpha$  as the  $\alpha$ -quantile of  $\{\tilde{V}_{i,b,T}, i = 1, \dots, B_T\}$ .

The following theorems establish the consistency of  $\hat{c}_\alpha$  and  $\hat{c}'_\alpha$ :

**Theorem 4** (Critical Values for SN-QR by Resampling). *Suppose the assumptions of Theorems 1 and 2 hold,  $b/T \rightarrow 0, b \rightarrow \infty, T \rightarrow \infty$  and  $B_T \rightarrow \infty$ . Then  $\hat{c}_\alpha \rightarrow_p c_\alpha$ .*

**Theorem 5** (Critical Values for CN-QR by Resampling). *Suppose the assumptions of Theorems 1 and 2 hold,  $b/T \rightarrow 0, b \rightarrow \infty, T \rightarrow \infty, B_T \rightarrow \infty$ , and  $\hat{A}_b$  is such that  $\hat{A}_b/A_b \rightarrow 1$ . Then  $\hat{c}'_\alpha \rightarrow_p c'_\alpha$ .*

**Comment 4.1.** Our subsampling method based on CN-QR or SN-QR produces consistent critical values in the regression case, and may also be of independent interest in the non-regression case. Our method differs from conventional subsampling in several respects. First, conventional subsampling uses fixed normalizations  $A_T$  or their consistent estimates. In contrast, in the case of SN-QR we use the random normalization  $\mathcal{A}_T$ , thus avoiding estimation of  $A_T$ . Second, conventional subsampling recenters by the full sample estimate  $\hat{\beta}(\tau_T)$ . Recentring in this way requires  $A_b/A_T \rightarrow 0$  for obtaining consistency (see Theorem 2.2.1 in Politis, Romano, and Wolf (1999)), but here we have  $A_b/A_T \rightarrow \infty$  when  $\xi > 0$ . Thus, when  $\xi > 0$  the extreme order QR statistics  $\hat{\beta}(\tau_T)$  diverge when  $\xi > 0$ , and the conventional subsampling is inconsistent. In contrast, to overcome the inconsistency, our approach instead uses  $\hat{\beta}(\tau_b)$  for recentring. This statistic itself may diverge, but because it is an intermediate order QR statistic, the speed of its divergence is strictly slower than that of  $A_T$ . Hence our method of recentring exploits the special structure of order statistics in both the regression and non-regression cases.

**4.2. Analytical Estimation of Critical Values.** Analytical inference uses the quantiles of the limit distributions found in Theorem 1. This approach is much more demanding in practice than the previous subsampling method.<sup>12</sup>

<sup>12</sup>The method developed below is also of independent interest in situations where the limit distributions involve Poisson processes with unknown nuisance parameters, as, for example, in Chernozhukov and Hong (2004).

Define the following random vector:

$$\widehat{Z}_\infty^*(k) = \widehat{\chi} \cdot \arg \min_{z \in \mathbb{R}^d} \left[ -k \bar{X}'_T(z + k^{-\widehat{\xi}} \widehat{\gamma}) + \sum_{t=1}^{\infty} \{ \mathcal{X}'_t(z + k^{-\widehat{\xi}} \widehat{\gamma}) - \widehat{\chi} \cdot \Gamma_t^{-\widehat{\xi}} \cdot \mathcal{X}'_t \widehat{\gamma} \}_+ \right], \quad (4.18)$$

for some consistent estimates  $\widehat{\xi}$  and  $\widehat{\gamma}$ , e.g., those given in equation (4.19); where  $\widehat{\chi} = 1$  if  $\widehat{\xi} < 0$  and  $\widehat{\chi} = -1$  if  $\widehat{\xi} > 0$ ,  $\{\Gamma_1, \Gamma_2, \dots\} = \{\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, \dots\}$ ;  $\{\mathcal{E}_1, \mathcal{E}_2, \dots\}$  is an i.i.d. sequence of standard exponential variables;  $\{\mathcal{X}_1, \mathcal{X}_2, \dots\}$  is an i.i.d. sequence with distribution function  $\widehat{F}_X$ , where  $\widehat{F}_X$  is any smooth consistent estimate of  $F_X$ , e.g., a smoothed empirical distribution function of the sample  $\{X_i, i = 1, \dots, T\}$ .<sup>13</sup> Moreover, the sequence  $\{\mathcal{X}_1, \mathcal{X}_2, \dots\}$  is independent from  $\{\mathcal{E}_1, \mathcal{E}_2, \dots\}$ . Also, let  $Z_\infty^*(k) = \sqrt{k} \widehat{Z}_\infty^*(k) / [\bar{X}'_T(\widehat{Z}_\infty^*(mk) - \widehat{Z}_\infty^*(k)) + \widehat{\chi}(m^{-\widehat{\xi}} - 1)k^{-\widehat{\xi}}]$ . The estimates  $\widehat{\mathcal{C}}_\alpha$  and  $\widehat{c}_\alpha$  are obtained by taking  $\alpha$ -quantiles of the variables  $\psi' \widehat{Z}_\infty^*(k)$  and  $\psi' Z_\infty^*(k)$ , respectively. In practice, these quantiles can only be evaluated numerically as described below.

The analytical inference procedure requires consistent estimators of  $\xi$  and  $\gamma$ . Theorem 4.5 of Chernozhukov (2005) provides the following estimators based on Pickands-type procedures:

$$\widehat{\xi} = \frac{-1}{\ln 2} \ln \frac{\bar{X}'_T(\widehat{\beta}(4\tau_T) - \widehat{\beta}(\tau_T))}{\bar{X}'_T(\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T))} \text{ and } \widehat{\gamma} = \frac{\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T)}{\bar{X}'_T(\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T))}, \quad (4.19)$$

which is consistent if  $\tau_T T \rightarrow \infty$  and  $\tau_T \rightarrow 0$ .

**Theorem 6** (Critical Values for SN-QR by Analytical Method). *Assume the conditions of Theorem 1 hold. Then for any estimators of the nuisance parameters such that  $\widehat{\xi} \rightarrow_p \xi$  and  $\widehat{\gamma} \rightarrow_p \gamma$ , we have that  $\widehat{\mathcal{C}}_\alpha \rightarrow_p c_\alpha$ .*

**Theorem 7** (Critical Values for CN-QR by Analytical Method). *Assume the conditions of Theorem 1 hold. Then, for any estimators of the nuisance parameters such that  $\widehat{\xi} \rightarrow_p \xi$  and  $\widehat{\gamma} \rightarrow_p \gamma$ , we have that  $\widehat{\mathcal{C}}'_\alpha \rightarrow_p c'_\alpha$ .*

**Comment 4.2.** Since the distributions of  $\widehat{Z}_\infty(k)$  and  $Z_\infty(k)$  do not have closed form, except in very special cases,  $\widehat{\mathcal{C}}'_\alpha$  and  $\widehat{c}_\alpha$  can be obtained numerically via the following Monte Carlo procedure. First, for each  $i = 1, \dots, B$  compute  $\widehat{Z}_{i,\infty}^*(k)$  and  $Z_{i,\infty}^*(k)$  using formula (4.18) by simulation, where the infinite summation is truncated at some finite value  $M$ .

<sup>13</sup>We need smoothness of the distribution regressors  $\mathcal{X}_i$  to guarantee uniqueness of the solution of the optimization problem (4.18); a similar device is used by De Angelis, Hall, and Young (1993) in the context of Edgeworth expansion for median regression. The empirical distribution function (edf) of  $X_i$  is not suited for this purpose, since it assigns point masses to sample points. However, making random draws from the edf and adding small noise with variance that is inversely proportional to the sample size produces draws from a smoothed empirical distribution function which is uniformly consistent with respect to  $F_X$ .

Second, take  $\widehat{c}'_\alpha$  and  $\widehat{c}_\alpha$  as the sample  $\alpha$ -quantiles of the samples  $\{\psi' \widehat{Z}_{i,\infty}^*(k), i = 1, \dots, B\}$  and  $\{\psi' Z_{i,\infty}^*(k), i = 1, \dots, B\}$ , respectively. We have found in numerical experiments that choosing  $M \geq 200$  and  $B \geq 100$  provides accurate estimates.

## 5. EXTREME VALUE VS. NORMAL INFERENCE: COMPARISONS

**5.1. Properties of Confidence Intervals with Unknown Nuisance Parameters.** In this section we compare the inferential performance of normal and extremal confidence intervals (CI) using the model:  $Y_t = X_t' \beta + U_t$ ,  $t = 1, \dots, 500$ ,  $d = 7$ ,  $\beta_j = 1$  for  $j \in \{1, \dots, 7\}$ , where the disturbances  $\{U_t\}$  are i.i.d. and follow either (1) a  $t$  distribution with  $\nu \in \{1, 3, 30\}$  degrees of freedom, or (2) a Weibull distribution with the shape parameter  $\alpha \in \{1, 3, 30\}$ . These distributions have EV indexes  $\xi = 1/\nu \in \{1, 1/3, 1/30\}$  and  $\xi = -1/\alpha \in \{-1, -1/3, -1/30\}$ , respectively. Regressors are drawn with replacement from the empirical application in Section 6.1 in order to match a real situation as closely as possible.<sup>14</sup> The design of the first type corresponds to tail properties of financial data, including returns and trade volumes; and the design of the second type corresponds to tail properties of microeconomic data, including birthweights, wages, and bids. Figures 2 and 3 plot coverage properties of CIs for the intercept and one of the slope coefficients based on subsampling the SN-QR statistic with  $B_T = 200$  and  $b = 100$ , and on the normal inference method suggested by Powell (1986) with a Hall-Sheather type rule for the bandwidth suggested in Koenker (2005).<sup>15</sup> The figures are based on QR estimates for  $\tau \in \{.01, .05, .10, .25, .50\}$ , i.e.  $\tau T \in \{5, 25, 50, 125, 250\}$ .

When the disturbances follow  $t$  distributions, the extremal CIs have good coverage properties, whereas the normal CIs typically *undercover* their performance deteriorates in the degree of heavy-tailedness and improves in the index  $\tau T$ . In heavy-tailed cases ( $\xi \in \{1, 1/3\}$ ) the normal CIs substantially undercover for extreme quantiles, as might be expected from the fact that the normal distribution fails to capture the heavy tails of the actual distribution of the QR statistic. In the thin-tailed case ( $\xi = 1/30$ ), the normal CIs still undercover for extreme quantiles. The extremal CIs perform consistently better than normal CIs, giving coverages close to the nominal level of 90%.

When the disturbances follow Weibull distributions, extremal CIs continue to have good coverage properties, whereas normal CIs either *undercover* or *overcover*, and their performance deteriorates in the degree of heavy-tailedness and improves in the index  $\tau T$ . In

<sup>14</sup>These data as well as the Monte-Carlo programs are deposited at [www.mit.edu/vchern](http://www.mit.edu/vchern).

<sup>15</sup> The alternative options implemented in the statistical package R to obtain standard errors for the normal method give similar results. These results are available from the authors upon request.

heavy-tailed cases ( $\xi = -1$ ) the normal CIs strongly overcover, which results from the overdispersion of the normal distribution relative to the actual distribution of QR statistics. In the thin-tailed cases ( $\xi = -1/30$ ) the normal CIs undercover and their performance improves in the index  $\tau T$ . In all cases, extremal CIs perform better than normal CIs, giving coverage rates close to the nominal level of 90% even for central quantiles.

We also compare forecasting properties of ordinary QR estimators and median-bias-corrected QR estimators of the intercept and slope coefficients, using the median absolute deviation and median bias as measures of performance (other measures may not be well-defined). We find that the gains to bias-correcting appear to be very small, except in the finite-support case with disturbances that are heavy-tailed near the boundary. We do not report these results for the sake of brevity.

**5.2. Practicalities and Rules of Thumb.** Equipped with both simulation experiments and practical experience, we provide a simple rule-of-thumb for the application of extremal inference. Recall that the order of a sample  $\tau$ -quantile in the sample of size  $T$  is the number  $\tau T$  (rounded to the next integer). This order plays a crucial role in determining whether extremal inference or central inference should be applied. Indeed, the former requires  $\tau T \rightarrow k$  whereas the latter requires  $\tau T \rightarrow \infty$ . In the regression case, in addition to the number  $\tau T$ , we need to take into account the number of regressors. As an example, let us consider the case where all  $d$  regressors are indicators that equally divide the sample of size  $T$  into subsamples of size  $T/d$ . Then the QR statistic will be determined by sample quantiles of order  $\tau T/d$  in each of these  $d$  subsamples. We may therefore think of the number  $\tau T/d$  as being a dimension-adjusted order for QR. A common simple rule for the application of the normal law is that the sample size is greater than 30. This suggests we should use extremal inference whenever  $\tau T/d \lesssim 30$ . This simple rule may or may not be conservative. For example, when regressors are continuous, our computational experiments indicate that normal inference performs as well as extremal inference as soon as  $\tau T/d \gtrsim 15 - 20$ , which suggests using extremal inference when  $\tau T/d \lesssim 15 - 20$  for this case. On the other hand, if we have an indicator variable that picks out 2% of the entire sample, as in the birthweight application presented below, then the number of observations below the fitted quantile for this subsample will be  $\tau T/50$ , which motivates using extremal inference when  $\tau T/50 \lesssim 15 - 20$  for this case. This rule is far more conservative than the original simple rule. Overall, it seems prudent to use both extremal and normal inference methods in most cases, with the idea that the discrepancies between the two can indicate extreme situations. Indeed, note that our methods based on subsampling perform very well even in the non-extreme cases (see Figures 2 and 3).

## 6. EMPIRICAL EXAMPLES

**6.1. Extremal Risk of a Stock.** We consider the problem of finding factors that affect the value-at-risk of the Occidental Petroleum daily stock return, a problem that is interesting for both economic analysis and real-world risk management.<sup>16</sup> Our data set consists of 1,000 daily observations covering the period 1996-1998. The dependent variable  $Y_t$  is the daily return of the Occidental Petroleum stock and the regressors  $X_{1t}$ ,  $X_{2t}$ , and  $X_{3t}$  are the lagged return on the spot price of oil, the lagged one-day return of the Dow Jones Industrials index (market return), and the lagged own return  $Y_{t-1}$ , respectively. We use a flexible asymmetric linear specification where  $X_t = (1, X_{1t}^+, X_{1t}^-, X_{2t}^+, X_{2t}^-, X_{3t}^+, X_{3t}^-)$  with  $X_{jt}^+ = \max(X_{jt}, 0)$ ,  $X_{jt}^- = -\min(X_{jt}, 0)$  and  $j \in \{1, 2, 3\}$ .

We begin by stating overall estimation results for the basic predictive linear model. A detailed specification and goodness-of-fit analysis of this model has been given in Chernozhukov and Umantsev (2001), whereas here we focus on the extremal analysis in order to illustrate the new inferential tools. Figure 4 plots QR estimates  $\hat{\beta}(\tau) = (\hat{\beta}_j(\tau), j = 0, \dots, 7)$  along with 90% pointwise confidence intervals. We use both extremal CIs (solid lines) and normal CIs (dashed lines). Figures 5 and 6 plot bias-corrected QR estimates along with pointwise CIs for the lower and upper tails, respectively.

We focus the discussion on the impact of downward movements of the explanatory variables, namely  $X_{1t}^-$ ,  $X_{2t}^-$ , and  $X_{3t}^-$ , on the extreme risk, that is, on the low conditional/predicted quantiles of the stock return. The estimate of the coefficient on the negative spot price of oil,  $X_{1t}^-$ , is positive in the lower tail of the distribution and negative in the center, but it is not statistically significant at the 90% level. However, the extremal CIs indicate that the distribution of the QR statistic is asymmetric in the far left tail, hence the economic effect of the spot price of oil may potentially be quite strong. Thus, past drops in the spot price of oil potentially strongly decrease the extreme risk. The estimate of the coefficient on the negative market return,  $X_{2t}^-$ , is significantly negative in the far left tail but not in the center of the conditional distribution. From this we may conclude that the past market drops appear to significantly increase the extreme risk. The estimates of the coefficient on the negative lagged own return,  $X_{3t}^-$  are significantly negative in the lower half of the conditional distribution. We may conclude that past drops in own return significantly increase extreme and intermediate risks.

Finally, we compare the CIs produced by extremal inference and normal inference. This empirical example closely matches the Monte-Carlo experiment in the previous section

---

<sup>16</sup> See Christoffersen, Hahn, and Inoue (1999), Diebold, Schuermann, and Stroughair (2000), Chernozhukov and Umantsev (2001), and Engle and Manganelli (2004).

with heavy-tailed  $t(3)$  disturbances. From this experiment, we expect that in the empirical example normal CIs would understate the estimation uncertainty and would be considerably more narrow than extremal CIs in the tails. As shown in Figures 5 and 6, normal CIs are indeed much more narrow than extremal CIs at  $\tau < .15$  and  $\tau > .85$ .

**6.2. Extremal Birthweights.** We investigate the impact of various demographic characteristics and maternal behavior on extremely low quantiles of birthweights of live infants born in the United States to black mothers of ages between 18 and 45. We use the June 1997 Detailed Natality Data published by the National Center for Health Statistics. Previous studies by Abreveya (2001) and Koenker and Hallock (2001) used the same data set, but they focused the analysis on typical birthweights, in a range between 2000 and 4500 grams. In contrast, equipped with extremal inference, we now venture far into the tails and study extremely low birthweight quantiles, in the range between 250 and 1500 grams. Some of our findings differ sharply from previous results for typical non-extremal quantiles.

Our decision to focus the analysis on black mothers is motivated by Figure 7 which shows a troubling heavy tail of low birthweights for black mothers. We choose a linear specification similar to Koenker and Hallock (2001). The response variable is the birthweight recorded in grams. The set of covariates include: ‘Boy,’ an indicator of infant gender; ‘Married,’ an indicator of whether the mother was married or not; ‘No Prenatal,’ ‘Prenatal Second,’ and ‘Prenatal Third,’ indicator variables that divide the sample into 4 categories: mothers with no prenatal visit (less than 1% of the sample), mothers whose first prenatal visit was in the second trimester, and mothers whose first prenatal visit was in the third trimester (The baseline category is mothers with a first visit in the first trimester, which constitute 83% of the sample); ‘Smoker,’ an indicator of whether the mother smoked during pregnancy; ‘Cigarettes/Day,’ the mother’s reported average number of cigarettes smoked per day; ‘Education,’ a categorical variable taking a value of 0 if the mother had less than a high-school education, 1 if she completed high school education, 2 if she obtained some college education, and 3 if she graduated from college; ‘Age’ and ‘Age<sup>2</sup>,’ the mother’s age and the mother’s age squared, both in deviations from their sample means.<sup>17</sup> Thus the control group consists of mothers of average age who had their first prenatal visit during the first trimester, that have not completed high school, and who did not smoke. The intercept in the estimated quantile regression model will measure quantiles for this group, and will therefore be referred to as the centercept.

---

<sup>17</sup>We exclude variables related to mother’s weight gain during pregnancy because they might be simultaneously determined with the birth-weights.

Figures 8 and 9 report estimation results for extremal low quantiles and typical quantiles, respectively. These figures show point estimates, extremal 90% CIs, and normal 90% CIs. Note that the centercept in Figure 8 varies from 250 to about 1500 grams, indicating the approximate range of birthweights that our extremal analysis applies to. In what follows, we focus the discussion only on key covariates and on differences between extremal and central inference.

While the density of birthweights, shown in Figure 7, has a finite lower support point, it has little probability mass near the boundary. This points towards a situation similar to the Monte Carlo design with Weibull disturbances, where differences between central and extremal inference occur only sufficiently far in the tails. This is what we observe in this empirical example as well. For the most part, normal CIs tend to be at most 15 percent narrower than extremal CIs, with the exception of the coefficient on ‘No Prenatal’, for which normal CIs are twice as narrow as extremal CIs. Since only 1.9 percent of mothers had no prenatal care, the sample size used to estimate this coefficient is only 635, which suggests that the discrepancies between extremal CIs and central CIs for the coefficient on ‘No prenatal’ should occur only when  $\tau \lesssim 30/635 = 5\%$ . As Figure 9 shows, differences between extremal CIs and normal CIs arise mostly when  $\tau \lesssim 10\%$ .

The analysis of extremal birthweights, shown in Figure 8, reveals several departures from findings for typical birthweights in Figure 9. Most surprisingly, smoking appears to have no negative impact on extremal quantiles, whereas it has a strong negative effect on the typical quantiles. The lack of statistical significance in the tails could be due to selection, where only mothers confident of good outcomes smoke, or to smoking having little or no causal effect on very extreme outcomes. This finding motivates further analysis, possibly using data sets that enable instrumental variables strategies.

Prenatal medical care has a strong impact on extremal quantiles and relatively little impact on typical quantiles, especially in the middle of the distribution. In particular, the impacts of ‘Prenatal Second’ and ‘Prenatal Third’ in the tails are very strongly positive. These effects could be due to mothers confident of good outcomes choosing to have a late first prenatal visit. Alternatively, these effects could be due to a late first prenatal visit providing better means for improving birthweight outcomes. The extremal CIs for ‘No-prenatal’ includes values between 0 and  $-800$  grams, suggesting that the effect of ‘No-prenatal’ in the tails is definitely non-positive and may be strongly negative.

#### APPENDIX A. PROOF OF THEOREM 1

The proof will be given for the case when  $\xi < 0$ . The case with  $\xi > 0$  follows very similarly.

STEP 1. Recall that  $V_t = Y_t - X'_t \beta_e$  and consider the point process  $\widehat{\mathbf{N}}$  defined by  $\widehat{\mathbf{N}}(F) := \sum_{t=1}^T 1\{(A_T V_t, X_t) \in F\}$  for Borel subsets  $F$  of  $E := [0, +\infty) \times \mathbf{X}$ . The point process  $\widehat{\mathbf{N}}$  converges in law in the metric space of point measure  $M_p(E)$ , that is equipped with the metric induced by the topology of vague convergence. The limit process is a Poisson point process  $\mathbf{N}$  characterized by the mean intensity measure  $m_{\mathbf{N}}(F) := \int_F (x' \gamma)^{1/\xi} u^{-1/\xi} du dF_X(x)$ . Given this form of the mean intensity measure we can represent

$$\mathbf{N}(F) := \sum_{t=1}^{\infty} 1\{(J_t, \mathcal{X}_t) \in F\} \quad (\text{A.20})$$

for all Borel subsets  $F$  of  $E := [0, +\infty) \times \mathbf{X}$ , where  $J_t = (\mathcal{X}'_t \gamma) \cdot \Gamma_t^{-\xi}$ ,  $\Gamma_t = \mathcal{E}_1 + \dots + \mathcal{E}_t$ , for  $t \geq 1$ ,  $\{\mathcal{E}_t, t \geq 1\}$  is an i.i.d. sequence of standard exponential variables,  $\{\mathcal{X}_t, t \geq 1\}$  is an i.i.d. sequence from the distribution  $F_X$ . Note that when  $\xi > 0$  the same result and representation holds, except that we define  $J_i = -(\mathcal{X}'_i \gamma) \cdot \Gamma_i^{-\xi}$  (with a change of sign).

The convergence in law  $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$  follows from the following steps. First, for any set  $F$  defined as intersection of a bounded rectangle with  $E$ , we have (a)  $\lim_{T \rightarrow \infty} E\widehat{\mathbf{N}}(F) = m_{\mathbf{N}}(F)$ , which follows from the regular variation property of  $F_U$  and C1, and (b)  $\lim_{T \rightarrow \infty} P[\widehat{\mathbf{N}}(F) = 0] = e^{-m_{\mathbf{N}}(F)}$ , which follows by Meyer's (1973) theorem by the geometric strong mixing and by observing that  $T \sum_{j=2}^{\lfloor T/k \rfloor} P((A_T V_1, X_1) \in F, (A_T V_j, X_j) \in F) \leq O(T \lfloor T/k \rfloor P((A_T V_1, X_1) \in F)^2) = O(1/k)$  by C1 and C5. Consequently, (a) and (b) imply by Kallenberg's theorem (Resnick 1987) that  $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ , where  $\mathbf{N}$  is a Poisson point process  $\mathbf{N}$  with intensity measure  $m_{\mathbf{N}}$ .

STEP 2. Observe that  $\widetilde{Z}_T(k) := A_T(\widehat{\beta}(\tau) - \beta_e) = \arg \min_{z \in \mathbb{R}^d} \sum_{t=1}^T \rho_{\tau}(A_T V_t - X'_t z)$ . To see this define  $z := A_T(\beta - \beta_e)$ . Rearranging terms gives  $\sum_{t=1}^T \rho_{\tau}(A_T V_t - X'_t z) \equiv -\tau T \bar{X}'_T z - \sum_{t=1}^T 1(A_T V_t \leq X'_t z)(A_T V_t - X'_t z) + \sum_{t=1}^T \tau A_T V_t$ . Subtract  $\sum_{t=1}^T \tau A_T V_t$  that does not depend on  $z$  and does not affect optimization, and define

$$\widetilde{Q}_T(z, k) := -\tau T \bar{X}'_T z + \sum_{t=1}^T \ell(A_T V_t, X'_t z) = -\tau T \bar{X}'_T z + \int_E \ell(u, x' z) d\widehat{\mathbf{N}}(u, x),$$

where  $\ell(u, v) := 1(u \leq v)(v - u)$ . We have that  $\widetilde{Z}_T(k) = \arg \min_{z \in \mathbb{R}^d} \widetilde{Q}_T(z, k)$ .

Since  $\ell$  is continuous and vanishes outside a compact subset of  $E$ , the mapping  $N \mapsto \int_E \ell(u, x' z) dN(u, x)$ , which sends elements  $N$  of the metric space  $M_p(E)$ , to the real line, is continuous. Since  $\tau T \bar{X}_T \rightarrow_p kE[X]$  and  $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ , by the Continuous Mapping Theorem we conclude that the finite-dimensional limit of  $z \mapsto \widetilde{Q}_T(z, k)$  is given by

$$z \mapsto \widetilde{Q}_{\infty}(z, k) := -kE[X]'z + \int_E \ell(j, x' z) d\mathbf{N}(j, x) := -kE[X]'z + \sum_{t=1}^{\infty} \ell(J_t, \mathcal{X}'_t z).$$

Next we recall the Convexity Lemma of Geyer (1996) and Knight (1999), which states that if (i) a sequence of convex lower-semicontinuous function  $\widetilde{Q}_T : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$  converges in distribution in the finite-dimensional sense to  $\widetilde{Q}_{\infty} : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$  over a dense subset of  $\mathbb{R}^d$ , (ii)  $\widetilde{Q}_{\infty}$  is finite over a



non-empty open set  $Z_0 \subset \mathbb{R}^d$ , and (iii)  $\tilde{Q}_\infty$  is uniquely minimized at a random vector  $\tilde{Z}_\infty$ , then any argmin of  $\tilde{Q}_T$ , denoted  $\tilde{Z}_T$ , converges in distribution to  $\tilde{Z}_\infty$ .

By the Convexity lemma we conclude that  $\tilde{Z}_T(k) \in \arg \min_{z \in \mathbb{R}^d} \tilde{Q}_T(z, k)$  converges in distribution to  $\tilde{Z}_\infty(k) = \arg \min_{z \in \mathbb{R}^d} \tilde{Q}_\infty(z, k)$ , where the random vector  $\tilde{Z}_\infty(k)$  is uniquely defined by Lemma 1 in Appendix E.

STEP 3. By C1,  $A_T(\beta(\tau) - \beta_e) \rightarrow k^{-\xi}\gamma$  as  $\tau T \rightarrow k$  and  $T \rightarrow \infty$ . Thus  $A_T(\hat{\beta}(\tau) - \beta(\tau)) \rightarrow_d \hat{Z}_\infty(k) := \tilde{Z}_\infty(k) + k^{-\xi}\gamma$ . Then

$$\hat{Z}_\infty(k) = \tilde{Z}_\infty(k) + k^{-\xi}\gamma = \arg \min_{z \in \mathbb{R}^d} [-kE[X]'(z + k^{-\xi}\gamma) + \sum_{t=1}^{\infty} \ell(J_t, \mathcal{X}'_t(z + k^{-\xi}\gamma))].$$

STEP 4. Similarly to step 2 it follows that

$$(\tilde{Z}_T(mk), \tilde{Z}_T(k)) \in \arg \min_{(z'_1, z'_2)' \in \mathbb{R}^{2d}} \tilde{Q}_T(z_1, mk) + \tilde{Q}_T(z_2, k)$$

weakly converges to

$$(\tilde{Z}_\infty(mk), \tilde{Z}_\infty(k)) = \arg \min_{(z'_1, z'_2)' \in \mathbb{R}^{2d}} \tilde{Q}_\infty(z_1, mk) + \tilde{Q}_\infty(z_2, k),$$

where the random vectors  $\tilde{Z}_\infty(k)$  and  $\tilde{Z}_\infty(mk)$  are uniquely defined by Lemma 1 in Appendix E.

Therefore it follows that

$$\left( \hat{Z}_T(k), \frac{\mathcal{A}_T}{A_T} \right) = \left( \hat{Z}_T(k), \frac{\bar{X}'_T(\tilde{Z}_T(km) - \tilde{Z}_T(k))}{\sqrt{\tau T}} \right) \rightarrow_d \left( \hat{Z}_\infty(k), \frac{E[X]'(\tilde{Z}_\infty(km) - \tilde{Z}_\infty(k))}{\sqrt{k}} \right).$$

By Lemma 1 in Appendix E,  $E[X]'(\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k)) \neq 0$  a.s., provided that  $mk - k > d$ . It follows by the Extended Continuous Mapping Theorem that

$$Z_T(k) = \frac{\mathcal{A}_T}{A_T} \hat{Z}_T(k) = \frac{\sqrt{\tau T} \hat{Z}_T(k)}{\bar{X}'_T(\tilde{Z}_T(mk) - \tilde{Z}_T(k))} \rightarrow_d Z_\infty(k) = \frac{\sqrt{k} \hat{Z}_\infty(k)}{E[X]'(\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k))}.$$

Using the relations  $\hat{Z}_\infty(k) = \tilde{Z}_\infty(k) + k^{-\xi}\gamma$  and  $\hat{Z}_\infty(mk) = \tilde{Z}_\infty(mk) + (mk)^{-\xi}\gamma$  and  $E[X]'\gamma = 1$  holding by C1, we can represent

$$Z_\infty(k) = \frac{\sqrt{k} \hat{Z}_\infty(k)}{E[X]'(\hat{Z}_\infty(mk) - \hat{Z}_\infty(k)) + (m^{-\xi} - 1)k^{-\xi}}. \quad \square$$

## APPENDIX B. PROOF OF THEOREM 2 AND 3

The results follows by Theorem 1 and the definition of convergence in distribution.  $\square$

## APPENDIX C. PROOF OF THEOREM 4 AND 5

We will prove Theorem 4. The proof of Theorem 5 follows similarly. The main step of the proof, step 1, is specific to our problem. Let  $G_T(x) := \Pr\{V_T \leq x\}$  and  $G(x) := \Pr\{V_\infty \leq x\} =$

$\limsup_{T \rightarrow \infty} G_T(x)$ .

STEP 1. Letting  $V_{i,b,T} := \mathcal{A}_{i,b,T} \psi'(\widehat{\beta}_{i,b,T}(\tau_b) - \beta(\tau_b))$ , define

$$\begin{aligned}\widehat{G}_{b,T}(x) &:= B_T^{-1} \sum_{i=1}^{B_T} 1\{\widehat{V}_{i,b,T} \leq x\} = B_T^{-1} \sum_{i=1}^{B_T} 1\left\{V_{i,b,T} + \mathcal{A}_{i,b,T} \psi'(\beta(\tau_b) - \widehat{\beta}(\tau_b)) \leq x\right\}, \\ \dot{G}_{b,T}(x; \Delta) &:= B_T^{-1} \sum_{i=1}^{B_T} 1\{V_{i,b,T} + (\mathcal{A}_{i,b,T}/A_b) \times \Delta \leq x\},\end{aligned}$$

where  $A_b = 1/Q_U(1/b)$  is the canonical normalizing constant. Then

$$1[V_{i,b,T} \leq x - \mathcal{A}_{i,b,T} w_T/A_b] \leq 1[\widehat{V}_{i,b,T} \leq x] \leq 1[V_{i,b,T} \leq x + \mathcal{A}_{i,b,T} w_T/A_b]$$

for all  $i = 1, \dots, B_T$ , where  $w_T = |A_b \psi'(\beta(\tau_b) - \widehat{\beta}(\tau_b))|$ .

The principal claim is that, under conditions of Theorem 3,  $w_T = o_p(1)$ . The claim follows by noting that for  $k_T = \tau_T T \rightarrow k > 0$  as  $b/T \rightarrow 0$  and  $T \rightarrow \infty$ ,

$$\begin{aligned}A_b \times (\beta(\tau_b) - \widehat{\beta}(\tau_b)) &\sim \frac{2^{-\xi} - 1}{Q_U(2k_T/b) - Q_U(k_T/b)} \times O_p\left(\frac{Q_U(2k_T/b) - Q_U(k_T/b)}{\sqrt{\tau_b \cdot T}}\right) \\ &= O_p\left(\frac{1}{\sqrt{\tau_b \cdot T}}\right) = O_p\left(\sqrt{\frac{b}{k_T T}}\right) = O_p\left(\sqrt{\frac{b}{k_T}}\right) = o_p(1)\end{aligned}\tag{C.21}$$

The first relation in (C.21) follows from two facts: First, by definition  $A_b := 1/Q_U(1/b)$  and by the regular variation of  $Q_U$  at 0 with exponent  $-\xi$ , for any  $l \searrow 0$ ,  $Q_U(l)(2^{-\xi} - 1) \sim Q_U(2l) - Q_U(l)$ . Second, since  $\tau_b = k_T/b$  and since  $\tau_b \times T = (k_T/b) \times T \sim (k/b) \times T \rightarrow \infty$  at a polynomial speed in  $T$  by  $T/b \rightarrow \infty$  at a polynomial speed in  $T$  by assumption,  $\widehat{\beta}(\tau_b)$  is the intermediate order regression quantile computed using the *full sample* of size  $T$ , so that by Theorem 3 in Chernozhukov (2005)

$$\beta(\tau_b) - \widehat{\beta}(\tau_b) = O_p\left(\frac{Q_U(2k_T/b) - Q_U(k_T/b)}{\sqrt{\tau_b \cdot T}}\right).\tag{C.22}$$

Given that  $w_T = o_p(1)$ , for some sequence of constants  $\Delta_T \searrow 0$  as  $T \rightarrow \infty$  the following event occurs wp  $\rightarrow 1$ :

$$M_T = \left\{ \begin{array}{l} 1[V_{i,b,T} < x - \mathcal{A}_{i,b,T} \Delta_T/A_b] \leq 1[V_{i,b,T} < x - \mathcal{A}_{i,b,T} w_T/A_b] \\ \leq 1[\widehat{V}_{i,b,T} < x] \\ \leq 1[V_{i,b,T} < x + \mathcal{A}_{i,b,T} w_T/A_b] \\ \leq 1[V_{i,b,T} < x + \mathcal{A}_{i,b,T} \Delta_T/A_b], \\ \text{for all } i = 1, \dots, B_T. \end{array} \right\}.$$

Event  $M_T$  implies

$$\dot{G}_{b,T}(x; \Delta_T) \leq \widehat{G}_{b,T}(x) \leq \dot{G}_{b,T}(x; -\Delta_T).\tag{C.23}$$

STEP 2. In this part we show that at the continuity points of  $G(x)$ ,  $\dot{G}_{b,T}(x; \pm \Delta_T) \rightarrow_p G(x)$ . First, by non-replacement sampling

$$E[\dot{G}_{b,T}(x; \Delta_T)] = P[V_b - \mathcal{A}_b \Delta_T/A_b \leq x].\tag{C.24}$$

Second, at the continuity points of  $G(x)$

$$\lim_{T \rightarrow \infty} E[\dot{G}_{b,T}(x; \Delta_T)] = \lim_{b \rightarrow \infty} P[V_b - \mathcal{A}_b \Delta_T / A_b \leq x] = P[\psi' Z_\infty(k) \leq x] = G(x). \quad (\text{C.25})$$

The statement (C.25) follows because  $V_b - \frac{\mathcal{A}_b \Delta_T}{A_b} = V_b + o_p(1) \rightarrow_d \psi' Z_\infty(k)$ , since by Theorem 1  $V_b \rightarrow_d \psi' Z_\infty(k)$  and by the proof of Theorem 1 and by  $\Delta_T \searrow 0$

$$\frac{\mathcal{A}_b \Delta_T}{A_b} = O_p(1) \cdot \Delta_T = O_p(1) \cdot o(1) = o_p(1).$$

Third, because  $\dot{G}_{b,T}(x, \Delta_T)$  is a U-statistic of degree  $b$ , by the LLN for U-statistics in Politis, Romano, and Wolf (1999),  $\text{Var}(\dot{G}_{b,T}(x, \Delta_T)) = o(1)$ . This shows that  $\dot{G}_{b,T}(x; \Delta_T) \rightarrow_p G(x)$ . By the same argument  $\dot{G}_{b,T}(x; -\Delta_T) \rightarrow_p G(x)$ .

STEP 3. Finally, since event  $M_T$  occurs wp  $\rightarrow 1$  and so does (C.23), by Step 2 it follows that  $\hat{G}_{b,T}(x) \rightarrow_p G(x)$  for each  $x \in \mathbb{R}$ . Finally, convergence of distribution functions at continuity points, implies convergence of quantile functions at continuity points. Therefore, by the Extended Continuous Mapping Theorem,  $\hat{c}_\alpha = \hat{G}_{b,T}^{-1}(\alpha) \rightarrow_p c_\alpha = G^{-1}(\alpha)$ , provided  $G^{-1}(\alpha)$  is a continuity point of  $G(x)$ .  $\square$

#### APPENDIX D. PROOF OF THEOREMS 6 AND 7

We will prove Theorem 7; the proof of Theorem 6 follows similarly.

We prove the theorem by showing that the law of the limit variables is continuous in the underlying parameters, which implies the validity of the proposed procedure. This proof structure is similar to the one used in the parametric bootstrap proofs, with the complication that the limit distributions here are non-standard. The demonstration of continuity poses some difficulties, which we deal with by invoking epi-convergence arguments and exploiting the properties of the Poisson process (A.20). We also carry out the proof for the case with  $\xi < 0$ ; the proof for the case with  $\xi > 0$  is identical apart from a change in sign in the definition of the points of the Poisson process, as indicated in the proof of Theorem 1.

Let us first list the basic objects with which we will work:

1. The parameters are  $\xi \in (-\infty, 0)$ ,  $\gamma \in \mathbb{R}^d$ , and  $F_X \in \mathcal{F}_X$ , a distribution function on  $\mathbb{R}^d$  with the compact support  $\mathbf{X}$ . We have the set of estimates such that:

$$\sup_{x \in \mathbf{X}} |F_X(x) - \hat{F}_X(x)| \rightarrow_p 0, \quad \hat{\xi} \rightarrow_p \xi, \quad \hat{\gamma} \rightarrow_p \gamma \quad \text{as } T \rightarrow \infty, \quad (\text{D.26})$$

where  $\hat{F}_X \in \mathcal{F}_X$ . The set  $\mathcal{F}_X$  is the set of non-lattice distributions defined in Appendix E. The underlying probability space  $(\Omega, \mathcal{F}, P)$  is the original probability space induced by the data.

2.  $\mathbf{N}$  is a Poisson random measure (PRM), with mean intensity measure  $m_{\mathbf{N}}$ , and points representable as:  $(\Gamma_j^{-\xi} \cdot \mathcal{X}'_j \gamma, \quad \mathcal{X}_j), \quad j = 1, 2, 3, \dots$   $\mathbf{N}$  is a random element of a complete and separable metric space of point measures  $(M_p(E), \rho_v)$  with metric  $\rho_v$  generated by the topology of vague convergence. The underlying probability space  $(\Omega', \mathcal{F}', P')$  is the one induced by Monte-Carlo draws of

points of  $\mathbf{N}$ . This law of  $\mathbf{N}$  in  $(M_p(E), \rho_p)$  will be denoted as  $\mathcal{L}(\mathbf{N}|\xi, \gamma, F_X)$ . The law depends only on the parameters  $(\xi, \gamma, F_X)$  of the intensity measure  $m_{\mathbf{N}}$ .

3. The random objective function (ROF) takes the form

$$\begin{aligned} z \mapsto \widehat{Q}_\infty(z; k) &= -kE[X]'(z + k^{-\xi}\gamma) + \int_E \{x'(z + k^{-\xi}\gamma) - u^{-\xi} \cdot x'\gamma\}_+ d\mathbf{N}(u, x) \\ &= -kE[X]'(z + k^{-\xi}\gamma) + \sum_{t=1}^{\infty} \{\mathcal{X}'_t(z + k^{-\xi}\gamma) - \Gamma_t^{-\xi} \cdot \mathcal{X}'_t\gamma\}_+, \end{aligned} \quad (\text{D.27})$$

and is a random element of the metric space of proper lower-semi-continuous functions  $(LC(\mathbb{R}^d), \rho_e)$ , equipped with the metric  $\rho_e$  induced by the topology of *epi-convergence*. Geyer (1996) and Knight (1999) provide a detailed introduction to epi-convergence, with connections to convexity and stochastic equi-semicontinuity. Moreover, this function is convex in  $z$ , which is a very important property to what follows. The law of  $z \mapsto \widehat{Q}_\infty(z; k)$  in  $(LC(\mathbb{R}^d), \rho_e)$  will be denoted as  $\mathcal{L}(\widehat{Q}_\infty(\cdot; k)|\xi, \gamma, F_X)$ . This law depends only on the parameters  $(\xi, \gamma, F_X)$ .

4. The extremum statistic  $\widehat{Z}_\infty(k) = \arg \min_{z \in \mathbb{R}^d} \widehat{Q}_\infty(z; k)$  is a random element in the metric space  $\mathbb{R}^d$ , equipped with the usual Euclidian metric. The law of  $\widehat{Z}_\infty(k)$  in  $\mathbb{R}^d$  will be denoted as  $\mathcal{L}(\widehat{Z}_\infty(k)|\xi, \gamma, F_X)$ . This law depends only on the parameters  $(\xi, \gamma, F_X)$ .

Next we collect together several weak convergence properties of the key random elements, which are most pertinent to establishing the final result.

A. A sequence of PRM  $(\mathbf{N}^m, m = 1, 2, \dots)$  in  $(M_p(E), \rho_p)$  defined by the sequence of intensity measures  $m_{\mathbf{N}^m}$  with parameters  $(\xi^m, \gamma^m, F_X^m)$  converges weakly to a PRM  $\mathbf{N}$  with intensity measure  $m_{\mathbf{N}}$  with parameters  $(\xi, \gamma, F_X)$  if the law of the former converges to the law of the latter with respect to the Bounded-Lipschitz metric  $\rho_w$  (or any other metric that metrizes weak convergence):

$$\lim_{m \rightarrow \infty} \rho_w(\mathcal{L}(\mathbf{N}^m|\xi^m, \gamma^m, F_X^m), \mathcal{L}(\mathbf{N}|\xi, \gamma, F_X)) = 0. \quad (\text{D.28})$$

The weak convergence of PRMs is equivalent to pointwise convergence of their Laplace functionals:

$$\lim_{m \rightarrow \infty} \varphi(f; \mathbf{N}^m) = \varphi(f; \mathbf{N}), \quad \forall f \in C_K^+(E), \quad (\text{D.29})$$

where  $C_K^+(E)$  is the set of continuous positive functions  $f$  defined on the domain  $E$  and vanishing outside a compact subset of  $E$ . The Laplace functional is defined as and equal to:

$$\varphi(f; \mathbf{N}) := E \left[ e^{\int_E f(u, x) d\mathbf{N}(u, x)} \right] = e^{(- \int_E [1 - e^{-f(u, x)}] dm_{\mathbf{N}}(u, x))}. \quad (\text{D.30})$$

B. A sequence of ROFs  $\{\widehat{Q}_\infty^m(\cdot; k), m = 1, 2, 3, \dots\}$  defined by the sequence of parameters  $\{(\xi^m, \gamma^m, F_X^m), m = 1, 2, 3, \dots\}$  converges weakly to the ROF  $\widehat{Q}_\infty(\cdot; k)$  defined by parameters  $(\xi, \gamma, F_X)$  in the metric space  $(LC(\mathbb{R}^d), \rho_e)$ , if the law of the former converges to the law of the latter with respect to the Bounded Lipschitz metric  $\rho_w$  (or any other metric that metrizes weak convergence):

$$\lim_{m \rightarrow \infty} \rho_w(\mathcal{L}(\widehat{Q}_\infty^m(\cdot; k)|\xi^m, \gamma^m, F_X^m), \mathcal{L}(\widehat{Q}_\infty(\cdot; k)|\xi, \gamma, F_X)) = 0. \quad (\text{D.31})$$

Moreover, since the objective functions are convex in  $z$ , the above weak convergence is equivalent to the finite-dimensional weak convergence:

$$(\widehat{Q}_\infty^m(z_j; k), j = 1, \dots, J) \rightarrow_d (\widehat{Q}_\infty(z_j; k), j = 1, \dots, J) \quad \text{in } \mathbb{R}^J \quad (\text{D.32})$$

for any finite collection of points  $(z_j, j = 1, \dots, J)$ . The result, that finite-dimensional convergence in distribution of convex functions implies epi-convergence in distribution, is due to Geyer (1996) and Knight (1999). Thus, in order to check (D.31) we only need to check (D.32).

**C.** In turn, the weak convergence of objective functions  $\{\widehat{Q}_\infty^m(\cdot; k), m = 1, 2, 3, \dots\}$  to  $\{\widehat{Q}_\infty(z; k)$  in  $(LC(\mathbb{R}^d), \rho_e)$  implies that as  $m \rightarrow \infty$  the weak convergence of the corresponding argmins:  $\widehat{Z}_\infty^m(k) \rightarrow_d \widehat{Z}_\infty(k)$  in  $\mathbb{R}^d$ , that is,

$$\lim_{m \rightarrow \infty} \rho_w(\mathcal{L}(\widehat{Z}_\infty(k)|\xi^m, \gamma^m, F_X^m), \mathcal{L}(\widehat{Z}_\infty(k)|\xi, \gamma, F_X)) = 0. \quad (\text{D.33})$$

The proof is now done in two steps:

**I.** We would like to show that the law  $\mathcal{L}(\widehat{Z}_\infty(k)|\xi', \gamma', F_X')$  is continuous at  $(\xi', \gamma', F_X') = (\xi, \gamma, F_X)$  for each  $(\xi, \gamma, F_X)$  in the parameter space, that is, for any sequence  $(\xi^m, \gamma^m, F_X^m, m = 1, 2, \dots)$  such that

$$|\xi^m - \xi| \rightarrow 0, |\gamma^m - \gamma| \rightarrow 0, \sup_{x \in \mathbf{X}} |F_X^m(x) - F_X(x)| \rightarrow 0 \quad (\text{D.34})$$

with  $F_X^m \in \mathcal{F}_X$ , we have

$$\rho_w(\mathcal{L}(\widehat{Z}_\infty(k)|\xi^m, \gamma^m, F_X^m), \mathcal{L}(\widehat{Z}_\infty(k)|\xi, \gamma, F_X)) \rightarrow 0. \quad (\text{D.35})$$

**II.** Given this continuity property, as  $|\widehat{\xi} - \xi| \rightarrow_p 0, |\widehat{\gamma} - \gamma| \rightarrow_p 0, \sup_{x \in \mathbf{X}} |\widehat{F}_X(x) - F_X(x)| \rightarrow_p 0$ , we have by the Continuous Mapping Theorem

$$\rho_w(\mathcal{L}(\widehat{Z}_\infty(k)|\widehat{\xi}, \widehat{\gamma}, \widehat{F}_X), \mathcal{L}(\widehat{Z}_\infty(k)|\xi, \gamma, F_X)) \rightarrow_p 0. \quad (\text{D.36})$$

That is, the law  $\mathcal{L}(\widehat{Z}_\infty(k)|\widehat{\xi}, \widehat{\gamma}, \widehat{F}_X)$  generated by the Monte Carlo procedure consistently estimates the limit law  $\mathcal{L}(\widehat{Z}_\infty(k)|\xi, \gamma, F_X)$ , which is what we needed to prove, since this result implies that the convergence of respective distribution functions at the continuity points of the limit distribution. Convergence of the distribution functions at the continuity points of the limit distribution implies convergence of the respective quantiles to the quantiles of the limit distribution provided the latter are positioned at the continuity points of the limit distribution function.

Thus it only remains to show the key continuity step I. We have that

$$(\text{D.34}) \xrightarrow{(1)} (\text{D.29}) \xrightarrow{(2)} (\text{D.28}) \xrightarrow{(3)} (\text{D.32}) \xrightarrow{(4)} (\text{D.31}) \xrightarrow{(5)} (\text{D.33}) \Leftrightarrow (\text{D.35})$$

where (1) follows by direct calculations: for  $g(u, x) = 1 - e^{-f(u, x)}$  and any  $f \in C_K(E)$

$$|\varphi(f; \mathbf{N}^m) - \varphi(f; \mathbf{N})| \leq \varphi(f; \mathbf{N}) \left| \exp \left\{ \int_E g(u, x) [dm_{\mathbf{N}} - dm_{\mathbf{N}^m}] \right\} - 1 \right| \rightarrow 0,$$

as  $\int_E (g(u, x) [dm_{\mathbf{N}} - dm_{\mathbf{N}^m}]) \rightarrow 0$ , which follows from the definition of the measure  $m_{\mathbf{N}}$  stated earlier; (2) follows by the preceding discussion in Step A; (3) follows by the continuity of the mapping

$$\mathbf{N} \mapsto \int_E (x'(z + k^{-\xi}) - u^{-\xi} \cdot x' \gamma)_+ d\mathbf{N}(u, x)$$

from  $(M_p(E), \rho_v)$  to  $\mathbb{R}$ , as noted in the proof of Theorem 1; and (4) and (5) follow by the preceding discussion in Step C.  $\square$

#### APPENDIX E. UNIQUENESS AND CONTINUITY

Define  $k := \lim_{T \rightarrow \infty} \tau T$  and fix an  $m$  such that  $k(m-1) > d$  where  $d = \dim(X)$ .

Let  $\{\mathcal{X}_t, t \geq 1\}$  be an i.i.d. sequence from a distribution function  $F_X$  such that  $E[\mathcal{X}\mathcal{X}']$  is positive definite. Define  $G_j := (kE[X] - \sum_{t \leq j} \mathcal{X}_t)' \{[\mathcal{X}_{j+1} \dots \mathcal{X}_{j+d}']^{-1}$  if the matrix  $[\mathcal{X}_{j+1} \dots \mathcal{X}_{j+d}]'$  is invertible, and  $G := (\infty, \dots, \infty)$  otherwise. Denote by  $\mathcal{F}_X(k)$  the class of distributions  $F_X$  for which  $P_{F_X}\{G_j \in \partial(0, 1)^d\} = 0$  for all integer  $j \geq 0$ .

**Definition** (Non-Lattice Condition Given  $k$  and  $m$ ).  $F_X \in \mathcal{F}_X(k')$  for both  $k' = k$  and  $k' = mk$ .

Denote the class of all non-lattice distributions as  $\mathcal{F}_X = \mathcal{F}_X(k) \cap \mathcal{F}_X(mk)$ .

**Lemma 1.** *If  $F_X \in \mathcal{F}_X$ , then  $\hat{Z}_\infty(k)$  and  $Z_\infty(k)$  are uniquely defined random vectors. Moreover, for any  $\psi \neq 0$ ,  $\psi' \hat{Z}_\infty(k)$  and  $\psi' Z_\infty(k)$  have continuous distribution functions.*

**Comment E.1.** The non-lattice condition is an analog of Koenker and Bassett's (1978) condition for uniqueness of quantile regression in finite samples. This condition trivially holds if the nonconstant covariates  $\mathcal{X}_{-1t}$  are absolutely continuous. Uniqueness therefore holds generically in the sense that for a fixed  $k$  adding arbitrarily small absolutely continuous perturbations to  $\{\mathcal{X}_{-1t}\}$  ensures it.

**PROOF: STEP 1.** We have from Theorem 1 that  $\hat{Z}_\infty(k) = \tilde{Z}_\infty(k) + c$  for some constant  $c$ , where  $\tilde{Z}_\infty$  is defined in Step 2 of the proof of Theorem 1. Chernozhukov (2005) shows that a sufficient condition for tightness of possibly set-valued  $\tilde{Z}_\infty(k)$  is  $E[\mathcal{X}\mathcal{X}'] > 0$ . Taking tightness as given, conditions for uniqueness and continuity of  $\tilde{Z}_\infty(k)$  can be established. Define  $\mathcal{H}$  as the set of all  $d$ -element subsets of  $\mathbb{N} = \{1, 2, 3, \dots\}$ . For  $h \in \mathcal{H}$ , let  $\mathcal{X}(h)$  and  $J(h)$  be the matrix with rows  $\{\mathcal{X}_t, t \in h\}$ , and vector with elements  $\{J_t, t \in h\}$ , respectively, where  $J_t$  are defined in the proof of Theorem 1. Let  $\mathcal{H}^* = \{h \in \mathcal{H} : |\mathcal{X}(h)| \neq 0\}$ . Note that  $\mathcal{H}^*$  is non-empty a.s. by  $E[\mathcal{X}\mathcal{X}']$  positive definite and is countable. By the same argument as in the proof of Theorem 3.1. of Koenker and Bassett (1978) at least one element of  $\tilde{Z}_\infty(k)$  takes the form  $z_h = \mathcal{X}(h)^{-1} J(h)$  for some  $h \in \mathcal{H}^*$ , and must satisfy a sub-gradient condition:

$$\zeta_k(z_h) := (kE[X] - \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}'_t z_h) \mathcal{X}_t)' \mathcal{X}(h)^{-1} \in [0, 1]^d,$$

and the argmin is unique if and only if  $\zeta_k(z_h) \in \mathcal{D} = (0, 1)^d$ . By the same argument as in the proof of Theorem 3.4 in Koenker and Bassett (1978),  $z_h$  must obey

$$k - d \leq \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}'_t z_h) \leq k.$$

Then, uniqueness holds for a fixed  $k > 0$  if  $P(\exists h \in \mathcal{H}^* : \zeta_k(z_h) \in \partial\mathcal{D}) = 0$ . To show this is the case, define  $\mathcal{M}(j)$  as the set of all  $j$ -element subsets of  $\mathbb{N}$ , and define for  $\mu \in \mathcal{M}(j)$ ,  $G(\mu, h) := (kE[X] - \sum_{t \in \mu} \mathcal{X}_t)' \mathcal{X}(h)^{-1}$  if  $\mathcal{X}(h)$  is invertible, and  $G(\mu, h) := (\infty, \dots, \infty)$  otherwise. Now note that if  $P_{F_X}\{G(\mu, h) \in \partial\mathcal{D}\} = 0$  for any  $h \in \mathcal{H}$  and  $\mu \in \mathcal{M}(j)$  such that  $h \cap \mu = \emptyset$  and any integer  $j \geq 0$ , then

$$\begin{aligned} & P(\exists h \in \mathcal{H}^* : \zeta_k(z_h) \in \partial\mathcal{D}) \\ & \leq P\{G(\mu, h) \in \partial\mathcal{D}, \exists h \in \mathcal{H}, \exists \mu \in \mathcal{M}(j), \exists j \geq 0 : h \cap \mu = \emptyset, k - d \leq j \leq k\} \\ & \leq \sum_{(k-d) \vee 0 \leq j \leq k} \sum_{h \in \mathcal{H}} \sum_{\mu \in \mathcal{M}(j) : h \cap \mu = \emptyset} P_{F_X}\{G(\mu, h) \in \partial\mathcal{D}\} = 0, \end{aligned}$$

since the summation is taken over the countable set. Finally, by the i.i.d. assumption and  $h \cap \mu = \emptyset$ ,  $P_{F_X}\{G(\mu, h) \in \partial\mathcal{D}\} = P_{F_X}\{G_j \in \partial\mathcal{D}\}$ , where  $G_j$  is defined above. Therefore,  $P_{F_X}\{G_j \in \partial\mathcal{D}\} = 0$  for all integer  $j \geq 0$  is the condition that suffices for uniqueness.

STEP 2. Next want to show that the distribution function of  $\psi' \tilde{Z}_\infty(k)$  has no point masses, that is  $P\{\psi' \tilde{Z}_\infty(k) = x\} = 0$  for each  $x \in \mathbb{R}$  and each vector  $\psi \neq 0$ , which is the equivalent to showing continuity of  $x \mapsto P\{\psi' \tilde{Z}_\infty(k) \leq x\}$  at each  $x$ . Indeed, from above we have that  $\{\tilde{Z}_\infty(k) = \mathcal{X}(h)^{-1} J(h)\}$  for some  $h \in \mathcal{H}^*$  a.s., and  $P\{\psi' \mathcal{X}(h)^{-1} J(h) = x | \{\mathcal{X}_t, t \geq 1\}\} = 0$  for each  $h \in \mathcal{H}^*$  a.s., since  $J(h)$  is absolute continuous conditional on  $\{\mathcal{X}_t, t \geq 1\}$  and  $\psi' \mathcal{X}(h)^{-1} \neq 0$ . Therefore,

$$P\{\tilde{Z}_\infty(k) = z\} \leq E \left[ \sum_{h \in \mathcal{H}^*} P\{\psi' \mathcal{X}(h)^{-1} J(h) = z | \{\mathcal{X}_t, t \geq 1\}\} \right] = 0,$$

by countability of  $\mathcal{H}$  and the law of iterated expectations.

STEP 3. Next we want to show that the distribution function of  $\psi' Z_\infty(k)$  has no point masses, that is  $P\{\psi' Z_\infty(k) = x\} = 0$  for each  $x \in \mathbb{R}$  and each vector  $\psi \neq 0$ . We have that

$$Z_\infty(k) = \sqrt{k}(\tilde{Z}_\infty(k) + c)/(E[X]'(\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k)))$$

for some  $c \neq 0$ . From Steps 1 and 2 we have that solutions  $\tilde{Z}_\infty(mk)$  and  $\tilde{Z}_\infty(k)$  are a.s. unique and a.s. take the form  $\tilde{Z}_\infty(k) = z_{h_1} = \mathcal{X}(h_1)^{-1} J(h_1)$  and  $\tilde{Z}_\infty(mk) = z_{h_2} = \mathcal{X}(h_2)^{-1} J(h_2)$  for some  $(h_1, h_2) \in \mathcal{H}^* \times \mathcal{H}^*$ . Furthermore,  $mk - k > d$  implies  $h_1 \neq h_2$  and hence a.s.  $z_{h_1} \neq z_{h_2}$ . Indeed, to see this by Step 2 we must have the inequality

$$k - d \leq \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}'_t z_{h_1}) \leq k \quad \text{and} \quad mk - d \leq \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}'_t z_{h_2}) \leq mk,$$

so that  $h_1 = h_2$  implies  $mk - k \leq d$ . Furthermore, arguing similarly to (Bassett and Koenker 1982)'s Theorem 2.2., we observe that for  $h_1$  and  $h_2$  defined above

$$\begin{aligned}
& -kE[X]'z_{h_1} + \int_E \ell(u, x'z_{h_1})d\mathbf{N}(u, x) - (mk - k)E[X]'z_{h_2} \\
& < -kE[X]'z_{h_2} + \int_E \ell(u, x'z_{h_2})d\mathbf{N}(u, x) - (mk - k)E[X]'z_{h_2} \\
& = -mkE[X]'z_{h_2} + \int_E \ell(u, x'z_{h_2})d\mathbf{N}(u, x) \\
& < -mkE[X]'z_{h_1} + \int_E \ell(u, x'z_{h_1})d\mathbf{N}(u, x) \\
& < -kE[X]'z_{h_1} + \int_E \ell(u, x'z_{h_1})d\mathbf{N}(u, x) - (mk - k)E[X]'z_{h_1}.
\end{aligned}$$

Solving this inequality we obtain  $(mk - k)E[X]'(z_{h_2} - z_{h_1}) > 0$ . We conclude therefore that a.s.  $E[X]'(\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k)) = E[X]' \mathcal{X}(h_2)^{-1}J(h_2) - E[X]' \mathcal{X}(h_1)^{-1}J(h_1) > 0$ . Moreover, conditional on  $\{\mathcal{X}_t \geq 1\}$  we can show by a perturbation argument that  $h_1 \neq h_2$  must be such that  $E[X]'(\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k)) = c'_1 J(h_2) + c'_2 J(h_1 \setminus h_2)$  for some constant  $c_2 \neq 0$ . Let us denote by  $\mathcal{G}$  the set of all pairs  $h_1 \neq h_2$  in  $\mathcal{H}^* \times \mathcal{H}^*$  that obey these two conditions.

From step 1 and from  $E[X]'(\tilde{Z}_\infty(mk) - \tilde{Z}_\infty(k)) > 0$  a.s., it follows that  $Z_\infty(k)$  is a proper random variable. Furthermore, for any  $x \in \mathbb{R}$  a.s. for any  $(h_1, h_2) \in \mathcal{G}$  and  $S(h_1, h_2) = E[X]' \mathcal{X}(h_2)^{-1}J(h_2) - E[X]' \mathcal{X}(h_1)^{-1}J(h_1)$ ,  $P\{\psi'(\mathcal{X}(h_1)^{-1}J(h_1) + c)/S(h_1, h_2) = x \cap (h_1 \times h_2) \in \mathcal{G} | \{\mathcal{X}_t, t \geq 1\}\} = 0$ . The claim follows because, for any  $(h_1, h_2) \in \mathcal{G}$ ,  $\psi'(\mathcal{X}(h_1)^{-1}J(h_1) + c)/S(h_1, h_2)$  is absolutely continuous conditional on  $\{\mathcal{X}_t, t \geq 1\}$  by  $\psi' \mathcal{X}(h_1)^{-1}J(h_1)$  and  $S(h_1, h_2)$  being jointly absolutely continuous conditional on  $\{\mathcal{X}_t, t \geq 1\}$  and by the non-singularity of transformation  $(w, v) \mapsto [w + c]/v$  over region  $v > 0$ . Therefore, for any  $x \in \mathbb{R}$ ,  $P\{\psi'Z_\infty(k) = x\}$  is bounded above by

$$E \left[ \sum_{h_1 \neq h_2 \in \mathcal{H}^* \times \mathcal{H}^*} P\{\psi' \mathcal{X}(h_1)^{-1}J(h_1)/S(h_1, h_2) = z \cap (h_1 \times h_2) \in \mathcal{G} | \{\mathcal{X}_t, t \geq 1\}\} \right] = 0,$$

by countability of  $\mathcal{H}$  and the law of iterated expectations.

#### REFERENCES

- ABREVEYA, J. (2001): "The effects of Demographics and Maternal Behavior on the Distribution of birth outcomes," *Empirical Economics*, 26(1), 247–257.
- BASSETT, JR., G., AND R. KOENKER (1982): "An empirical quantile function for linear models with iid errors," *J. Amer. Statist. Assoc.*, 77(378), 407–415.
- BERTAIL, P., C. HAEFKE, D. POLITIS, AND H. WHITE (2004): "A subsampling approach to estimating the distribution of diverging extreme statistics with applications to assessing financial market risks," *J. Econometrics*.
- BICKEL, P., AND D. FREEDMAN (1981): "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics*, 9, 1196–1217.
- CABALLERO, R., AND E. ENGEL (1999): "Explaining Investment Dynamics in U.S. Manufacturing: a Generalized (S,s) Approach," *Econometrica*, 67(4), 783–826.
- CHERNOZHUKOV, V. (1998): "Nonparametric Extreme Regression Quantiles," 1998, Working Paper, Stanford. Presented at Princeton Econometrics Seminar in December 1998.
- (2005): "Extremal quantile regression," *Ann. Statist.*, 33(2), 806–839.



- CHERNOZHUKOV, V., AND H. HONG (2004): "Likelihood estimation and inference in a class of nonregular econometric models," *Econometrica*, 72(5), 1445–1480.
- CHERNOZHUKOV, V., AND L. UMANTSEV (2001): "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics*, 26(1), 271–293.
- CHRISTOFFERSEN, P. F., J. HAHN, AND A. INOUE (1999): "Testing, Comparing, and Combining Value-at-Risk Measures," SSRN working paper.
- DE ANGELIS, D., P. HALL, AND G. A. YOUNG (1993): "Analytical and bootstrap approximations to estimator distributions in  $L^1$  regression," *J. Amer. Statist. Assoc.*, 88(424), 1310–1316.
- DEKKERS, A., AND L. DE HAAN (1989): "On the estimation of the extreme-value index and large quantile estimation," *Ann. Statist.*, 17(4), 1795–1832.
- DIEBOLD, F., T. SCHUERMANN, AND J. STROUGHAIR (2000): "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management," *J. Risk Finance*, pp. 30–36.
- DONALD, S. G., AND H. J. PAARSCH (2002): "Superconsistent estimation and inference in structural econometric models using extreme order statistics," *J. Econometrics*, 109(2), 305–340.
- ENGLE, R. F., AND S. MANGANELLI (2004): "CAViaR: conditional autoregressive value at risk by regression quantiles," *J. Bus. Econom. Statist.*, 22(4), 367–381.
- FAMA, E. F. (1965): "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34–105.
- FEIGIN, P. D., AND S. I. RESNICK (1994): "Limit distributions for linear programming time series estimators," *Stochastic Process. Appl.*, 51(1), 135–165.
- FLINN, C., AND J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18, 115–190.
- GEYER, C. J. (1996): "On the asymptotics of convex stochastic optimization," Technical Report, University of Minnesota, Department of Statistics.
- GNEDENKO, B. (1943): "Sur la distribution limitée du terme d' une série aléatoire," *Ann. math.*, 44, 423–453.
- HIRANO, K., AND J. R. PORTER (2003): "Asymptotic efficiency in parametric structural models with parameter-dependent support," *Econometrica*, 71(5), 1307–1338.
- KIEFER, N. M., T. J. VOGELSANG, AND H. BUNZEL (2000): "Simple robust testing of regression hypotheses," *Econometrica*, 68(3), 695–714.
- KNIGHT, K. (1999): "Epi-convergence and Stochastic Equisemicontinuity," Technical Report, University of Toronto, Department of Statistics (<http://www.utstat.toronto.edu/keith/papers/>).
- (2001): "Limiting distributions of linear programming estimators," *Extremes*, 4(2), 87–103 (2002).
- KOENKER, R. (2005): *Quantile regression*, vol. 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.
- KOENKER, R., AND G. S. BASSETT (1978): "Regression quantiles," *Econometrica*, 46, 33–50.
- KOENKER, R., AND K. HALLOCK (2001): "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, forthcoming.
- LAPLACE, P.-S. (1818): *Théorie analytique des probabilités*. Éditions Jacques Gabay (1995), Paris.
- MANDELROT, M. (1963): "Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.
- MEYER, R. M. (1973): "A poisson-type limit theorem for mixing sequences of dependent 'rare' events," *Ann. Probability*, 1, 480–483.
- POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. Springer-Verlag, New York.
- PORTNOY, S., AND J. JUREČKOVÁ (1999): "On extreme regression quantiles," *Extremes*, 2(3), 227–243 (2000).
- RESNICK, S. (1987): *Extreme Values, Regular Variation, and Point Processes*. Springer.
- SEN, A. (1973): *On Economic Inequality*. Oxford University Press.
- TIMMER, C. P. (1971): "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency," *Journal of Political Economy*, 79, 776–794.
- ZIPF, G. (1949): *Human Behavior and the Principle of Last Effort*. Cambridge, MA: Addison-Wesley.

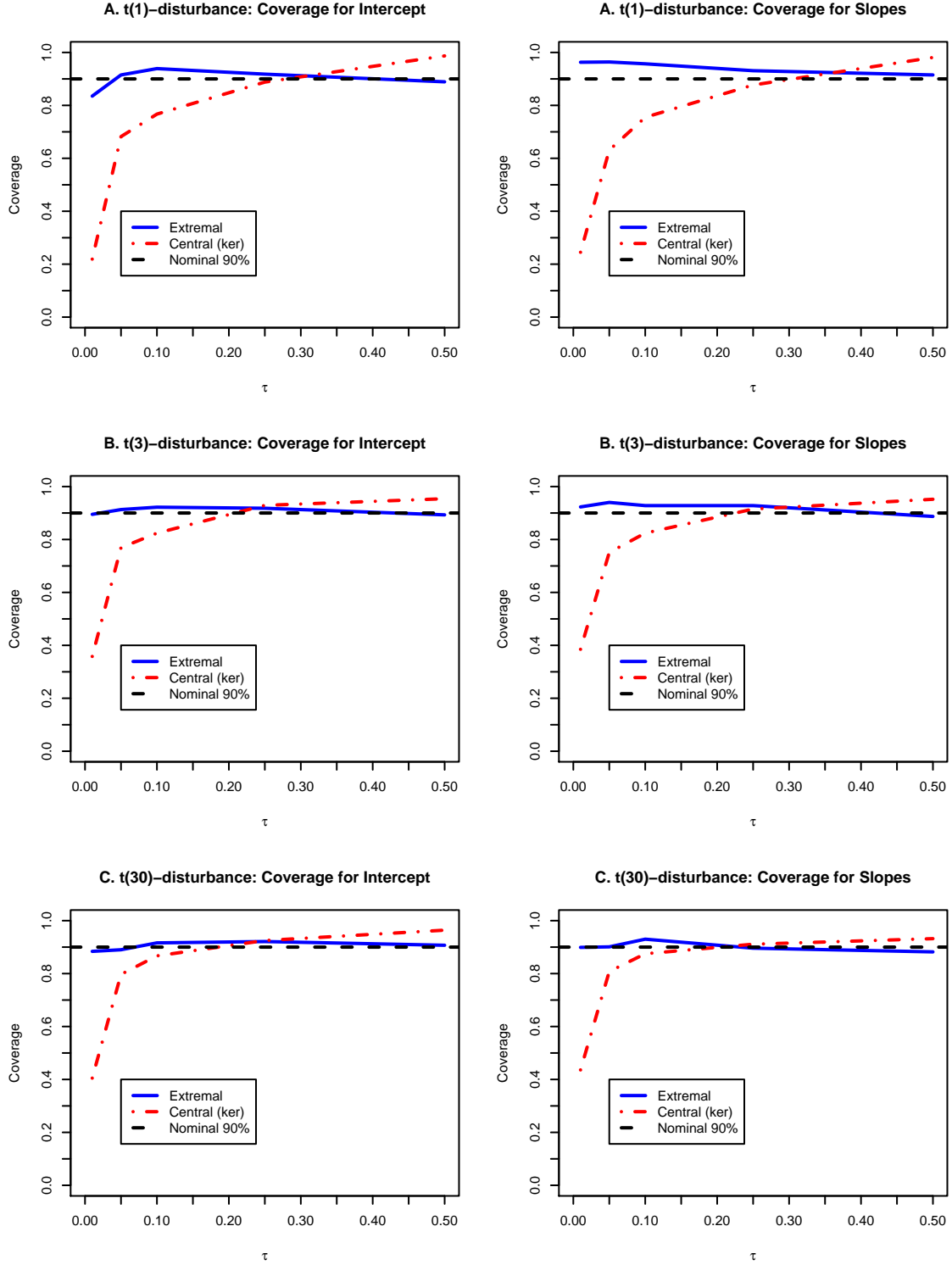


FIGURE 2. Coverage of extremal confidence intervals and normal confidence intervals when Disturbances are  $t(\nu)$ ,  $\nu \in \{1, 3, 30\}$ . Based on 1,000 repetitions.

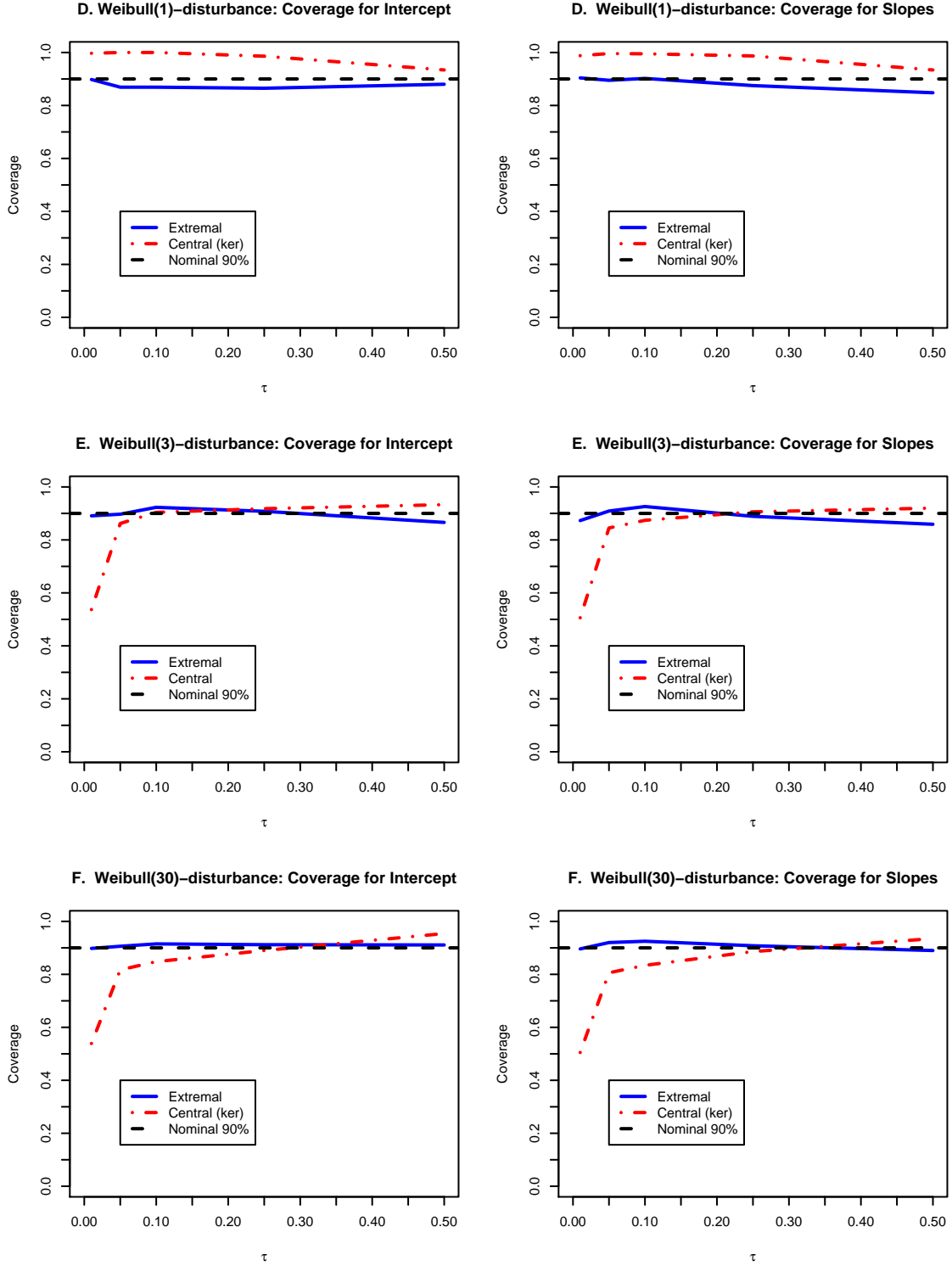


FIGURE 3. Coverage of extremal confidence intervals and normal confidence intervals when disturbances are Weibull  $(\alpha)$ ,  $\alpha \in \{1, 3, 30\}$ . Based on 1,000 repetitions.

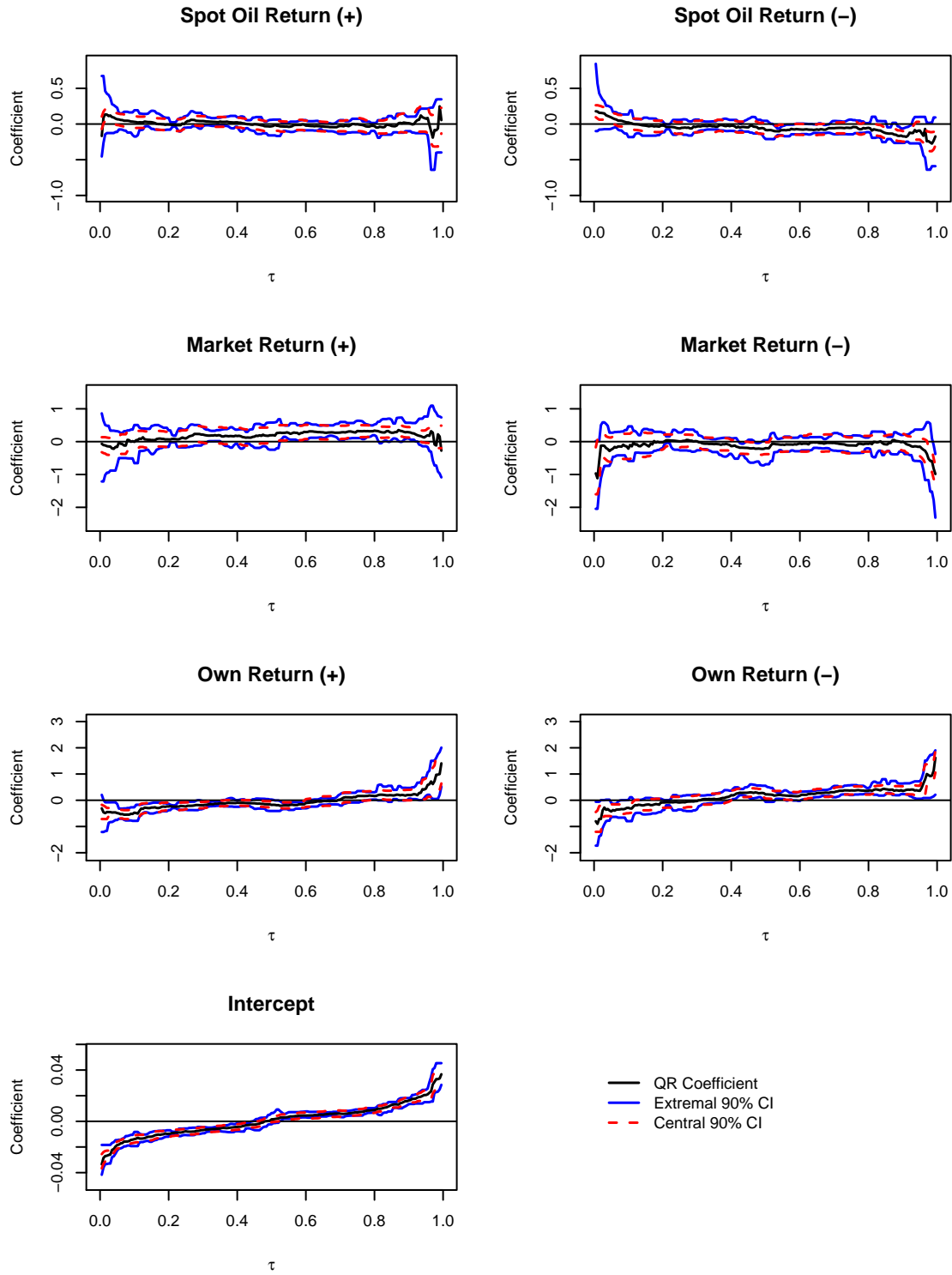


FIGURE 4. QR coefficient estimates and 90% pointwise confidence intervals. The solid lines depict extremal confidence intervals. The dashed lines depict normal confidence intervals.

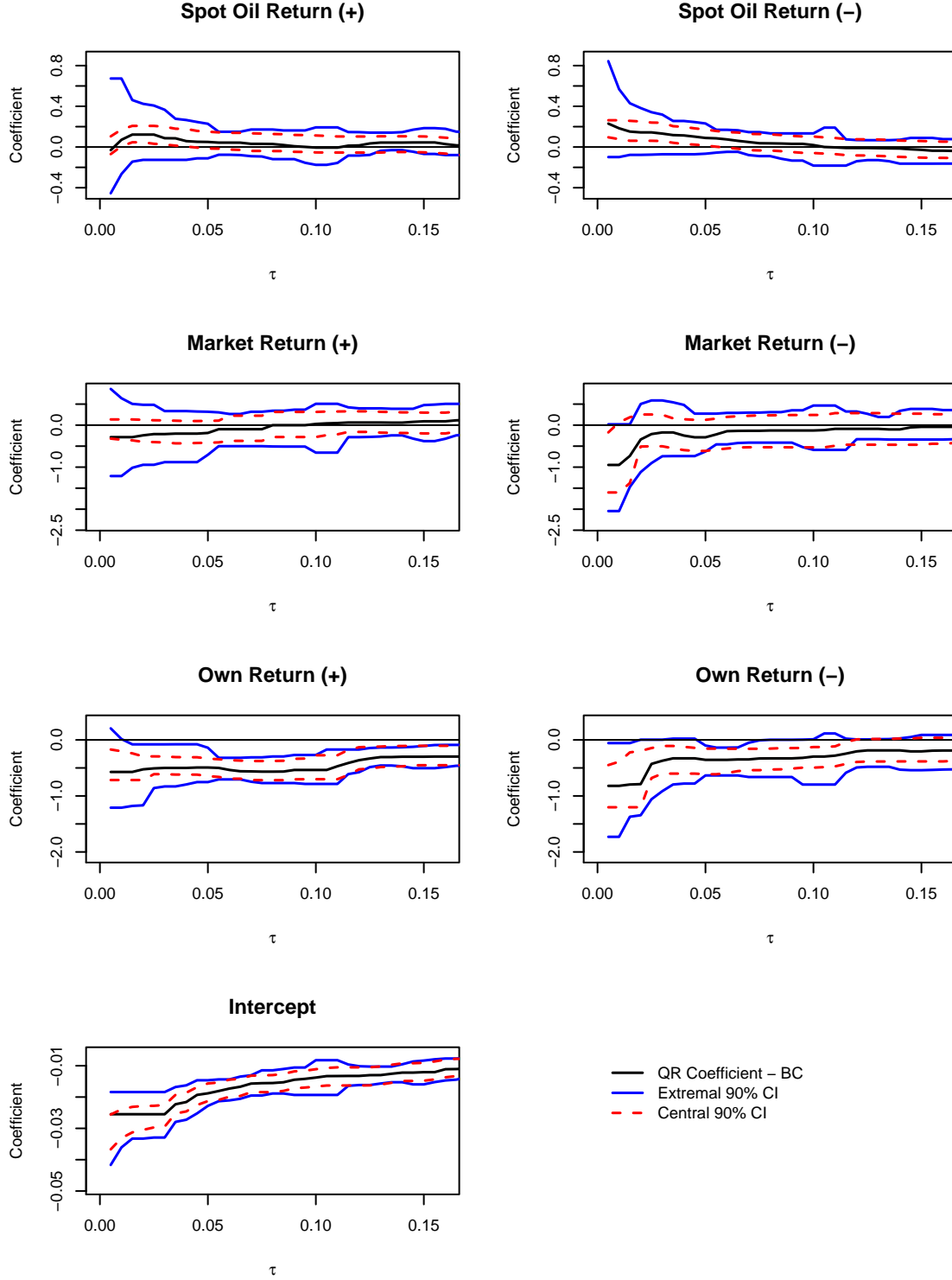


FIGURE 5. Bias-corrected QR coefficient estimates and 90% pointwise confidence intervals for  $\tau \leq .15$ . The solid lines depict extremal confidence intervals. The dashed lines depict normal confidence intervals.

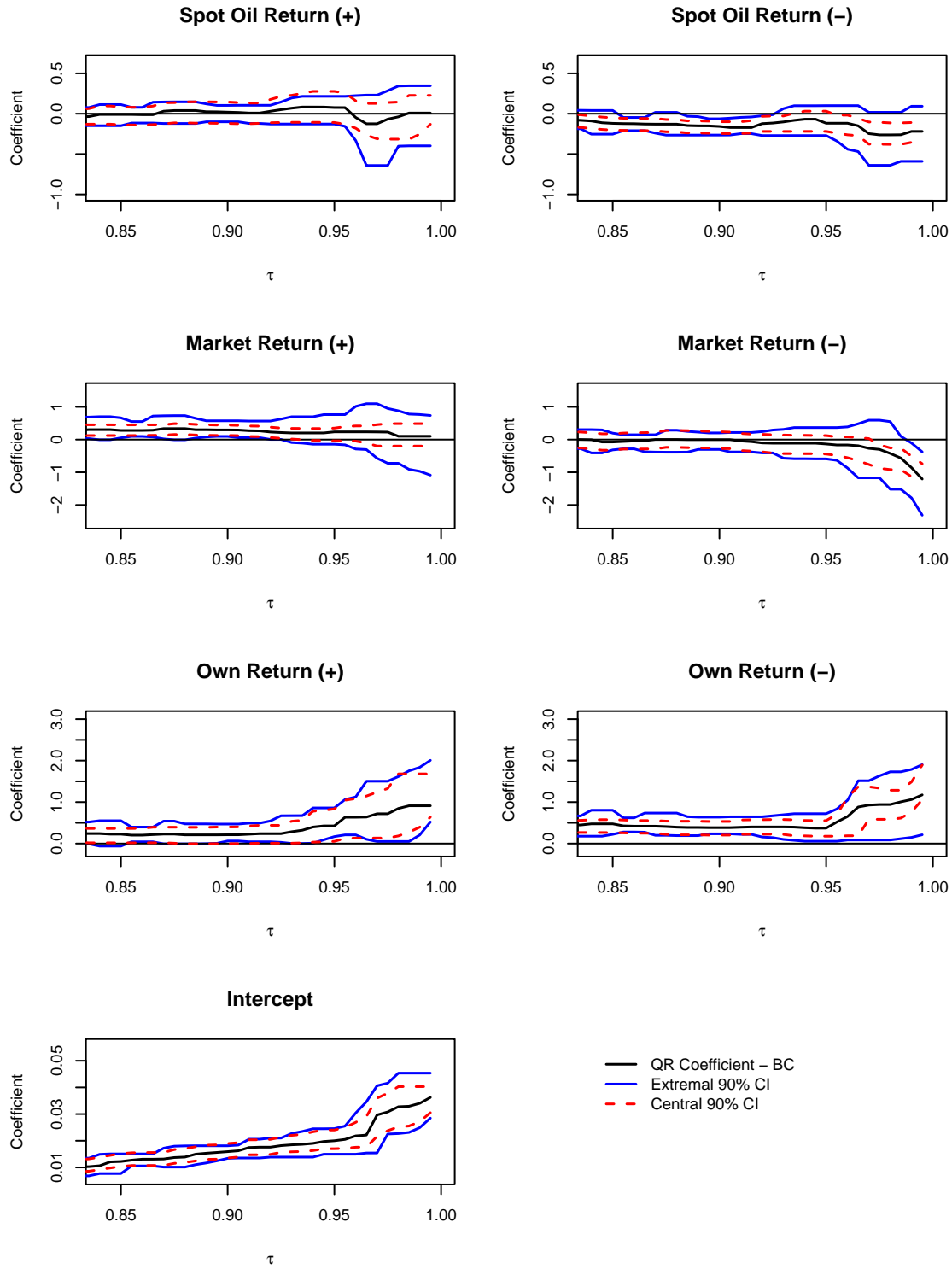


FIGURE 6. Bias-corrected QR coefficient estimates and 90% pointwise intervals for  $\tau \geq .85$ . The solid lines depict extremal confidence intervals. The dashed lines depict normal confidence intervals.

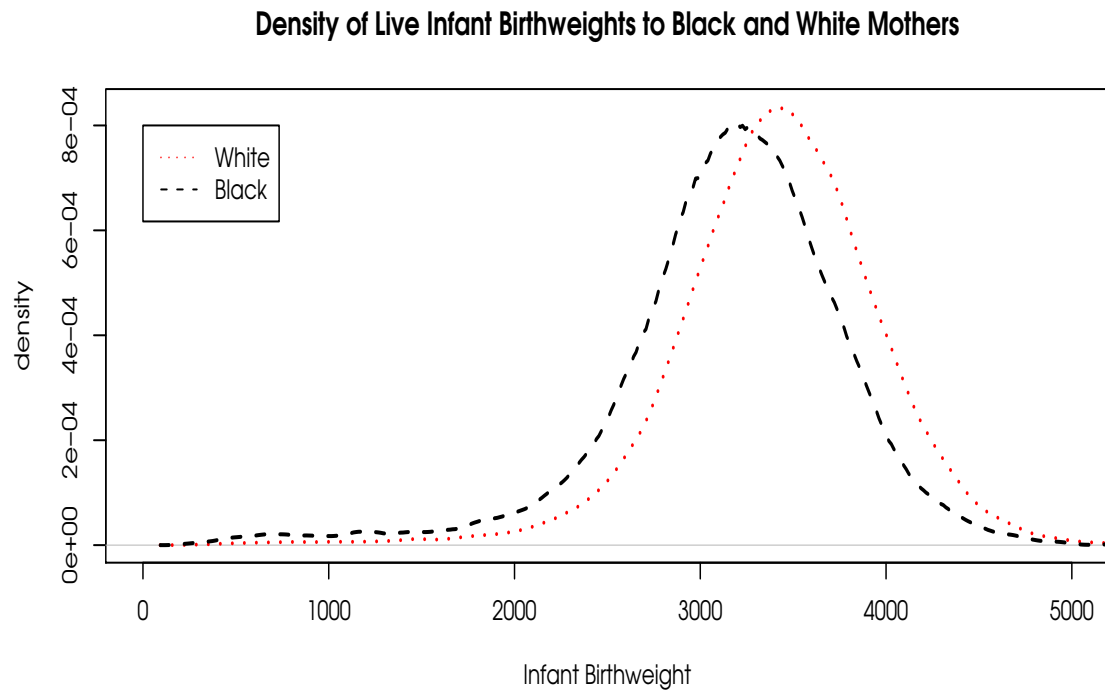


FIGURE 7. Birthweight densities for black and white mothers.

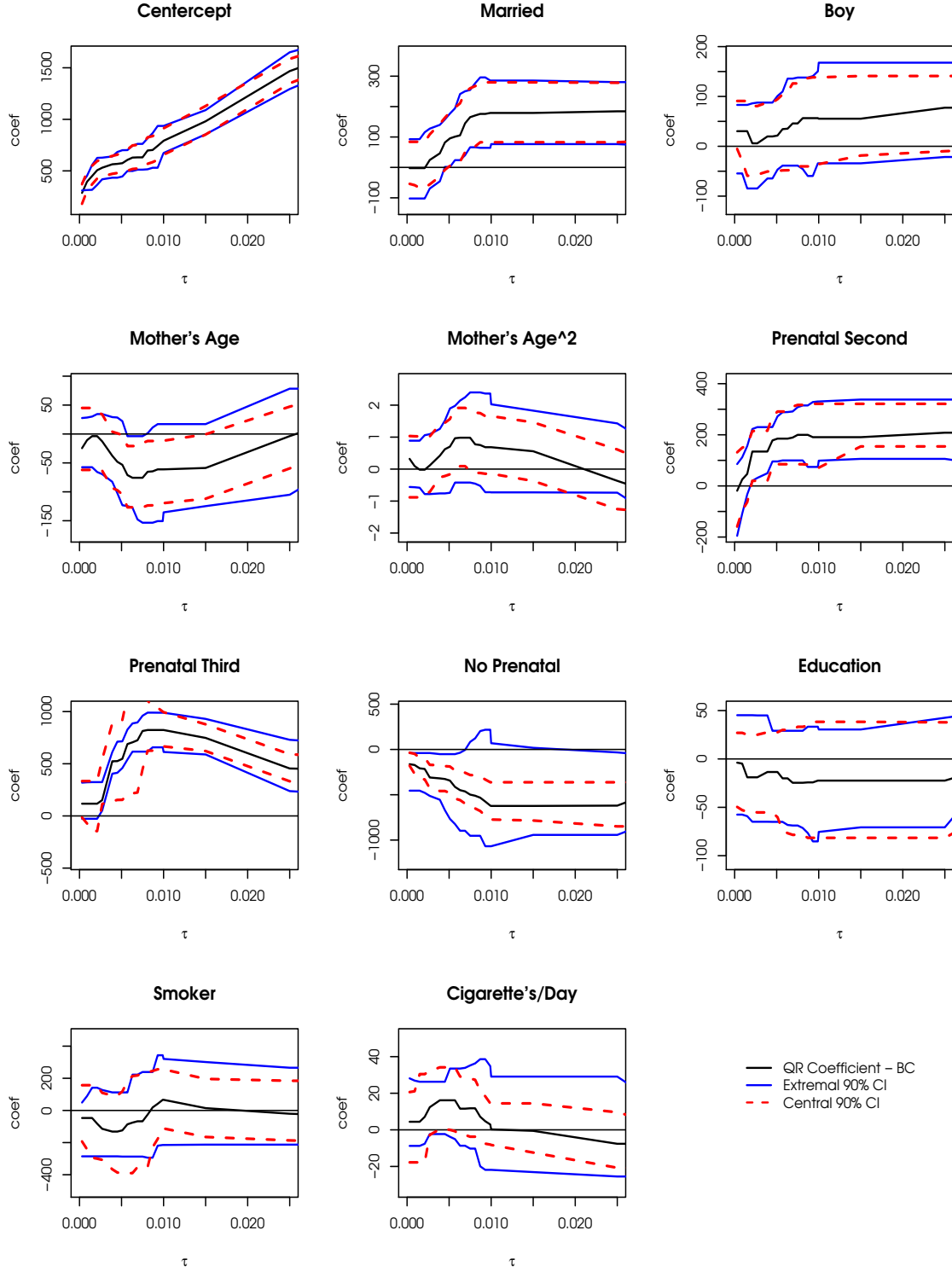


FIGURE 8. Bias-corrected QR coefficient estimates and 90% pointwise confidence intervals for  $\tau \leq .025$ . The solid lines depict extremal confidence intervals. The dashed lines depict normal confidence intervals.



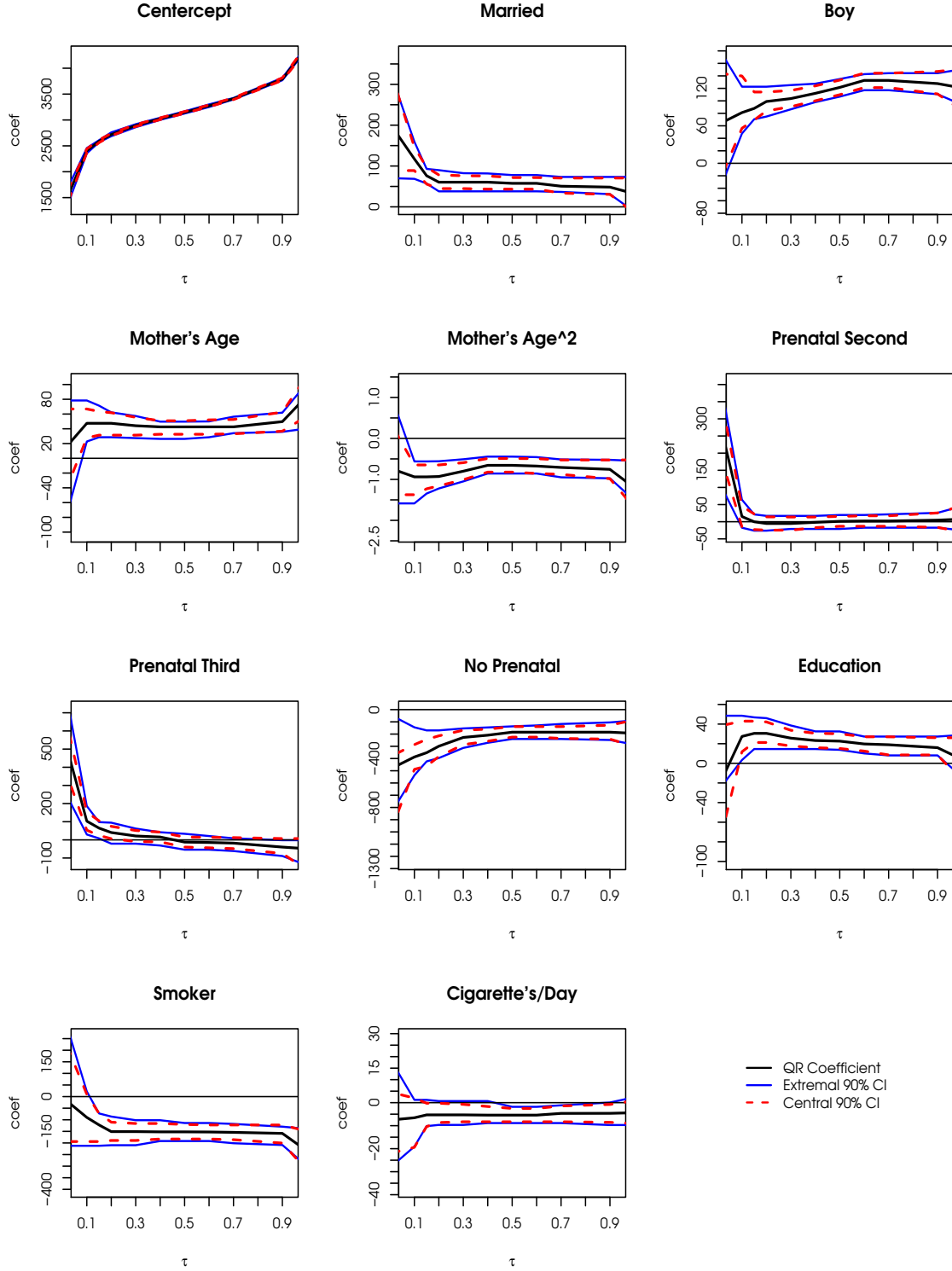


FIGURE 9. QR coefficient estimates and 90% pointwise confidence intervals for  $\tau \in [.025, .975]$ . The solid lines depict extremal confidence intervals. The dashed lines depict normal confidence intervals.