INFERENCE FOR EXTREMAL CONDITIONAL QUANTILE MODELS (EXTREME VALUE INFERENCE FOR QUANTILE REGRESSION)

VICTOR CHERNOZHUKOV

Abstract. Quantile regression is a basic tool for estimation of conditional quantiles of a response variable given a vector of regressors. It can be used to measure the effect of covariates not only in the center of a distribution, but also in the upper and lower tails. Quantile regression applied to the tails, or simply extremal quantile regression is of interest in numerous economic and financial applications. For example, it can be employed to measure conditional value-at-risk, production-efficiency, adjustment bands in the (S,s) models, and cost functions of most efficient bidders in auctions.

In order to facilitate the applications, this paper provides feasible inference tools that rely upon extreme value approximations of the distribution of self-normalized extremal quantile regression statistics. The methods are simple to implement in practice and are of independent interest even in the non-regression case. The value of the methods is explored in the analysis of extremely low percentiles of live infant birthweights (in the ranges between 250 and 1500 grams) and in the study of factors of extreme fluctuations of a stock return.

Key Words: Quantile Regression

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

Data and software in R available by request

Date: May, 2002, revised January, 2005. This is still a preliminary revision.

1. Introduction

This paper provides feasible inference methods for extremal quantile regression (QR) models, which are based on extreme value theory. The models have many vital applications to extremal phenomena arising in economics and other sciences, ranging from extreme risks to extreme birthweights. We provide feasible inference methods for these models and illustrate them with an analysis of economic determinants of extremely low birthweights, in the range between 250 and 1500 grams, and the analysis of economic factors of extreme fluctuations of a stock return.

The inferential methods of this paper are based on the extreme value theory for QR, namely on the sampling theory for QR that we develop specifically for "the tails." Formally the tail situation occurs when the quantile index $\tau \in (0, 1)$ is either small, close to zero, or large, close to 1. Without loss of generality we assume the former, and, by "close to 0" mean that $\tau T \to k > 0$ as the sample size $T \to \infty$. Under these conditions, the central limit theorem fails to hold, but different laws of extreme events arise. This approach, underlying also the classical extreme value theory in univariate settings, turns out to provide distributional approximations for QR that are of better quality in the tails than conventional normal approximations. However, there are fundamental difficulties that have previously made the inference for QR based on this approach both elusive and infeasible.

Main inference methods break down for several reasons. First, the basic inference approach using limit distributions of Wald statistics is not feasible, because extreme value approximations, both in the classical univariate case and in the regression case, rely on canonical normalizing constants used to achieve non-degenerate asymptotic distributions. Consistent estimation of these constants is not always feasible without additional strong assumptions. This is the first difficulty encountered here as well as in the classical univariate case (Bertail, Haefke, Politis, and White (2004)). Second, basic resampling methods also fail to deliver proper inference. The conventional bootstrap fails due to the nonstandard behavior of extremal quantile statistics (as shown by Bickel and Freedman (1981) in the univariate case). Conventional subsampling (Politis, Romano, and Wolf 1999) is also inconsistent, at least in in the unbounded support case where QR statistics diverge. Moreover, conventional subsampling suffers from the first difficulty as well, requiring consistent estimates of the canonical normalizing constants.

In this paper we develop two types of inference approaches that overcome these difficulties, a resampling approach and an analytical approach, with the former approach being the principal one. Both approaches are based on self-normalized QR (SN-QR) statistics that employ random normalization (instead of generally infeasible normalization by canonical constants). The use of SN-QR allows us to derive *feasible* limit distributions, which underlie either of our inference approaches. Moreover, our resampling approach is a modified subsampling method applied to SN-QR statistics. The approach *entirely avoids* not only estimating the canonical normalizing constants but also all other nuisance tail parameters known to be very hard to estimate. The modified subsampling method explores the special relationship between the rates of convergence of extremal and intermediate QR statistics, which allows for suitably estimating the centering constants in subsamples. This paper also provides inferential methods for canonically-normalized QR (CN-QR), for completeness, but also shows that much stronger assumptions are required for their feasibility.

This paper contributes to the existing literature by providing for the first time the general feasible inference for extremal quantile regression. This addresses a problem that was first asked in the context of estimating probabilistic frontier functions (e.g., Timmer (1971) and Aigner, Amemiya, and Poirier (1976).) The inferential methods proposed in this paper partially rely on limit results in Chernozhukov (2000b) who studied the limits of canonically-normalized QR under the extreme order condition $\tau T \to k > 0$. However, the limit theory in Chernozhukov (2000b) was more theoretical in nature and provided limit results that are infeasible for purposes of inference, in the sense explained earlier. Related limit results for the linear programming estimator, that corresponds to taking $\tau = \tau T = 0$, were considered by Feigin and Resnick (1994), Portnoy and Jurečková (1999), Knight (2001), and Chernozhukov (1998), all in different contexts and at various levels of generality. The linear programming estimator is most suited to the problem of estimating the finite boundary of the data, e.g. as in image processing and other technometric applications alike, whereas the current approach of taking $\tau T \to k > 0$ 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 and build upon this previous literature.

The paper is organized as follows. Section 2 motivates the analysis and formulates the econometric problem. Section 3 describes the model and regularity conditions. Section 4 describes the limit behavior of self-normalized QR statistics. Section 5 describes the inference methods. Section 6 describes the Monte-Carlo work. Section 7 presents empirical applications, and the Appendix collects proofs and figures.

2. Economic Problems and Econometric Problems

2.1. **Extremal Conditional Quantiles in Economic Analysis.**

There are many useful applications of inferential methods developed in this paper. For purposes of discussion in this section, let $Q_Y(\tau|X)$ denote the τ-quantile of a response variable Y given regressors X, that is the $\tau-$conditional quantile function.

Example 1. Determinants of Value-at-Risk. Value-at-risk analysis seeks to forecast or explain very low quantiles of an institution's portfolio return Y using today's available information W. This type of analysis is a basic, every day activity that banking and other financial institutions perform, as required by the Securities and Exchange Commission and the Basle Committee on Banking Supervision. The new inferential tools should be pertinent there, as the value-at-risk problem naturally lends itself to the quantile regression domain, e.g. Chernozhukov and Umantsev (2001) and Engle and Manganelli (2001). As a risk measure value-at-risk is motivated by the safety-first decision principle (Roy 1952), in which one makes optimal decisions subject to the constraint that a disastrous event occurs with a small probability. This and similar measures are often used in real-life financial management, insurance, and actuarial science (Embrechts, Klüppelberg, and Mikosch 1997).

Example 2. Determinants of Very Low Birthweights. In the analysis of infant birth weights, we may be interested in how smoking, absence of prenatal care, and other types of maternal behavior affect various quantiles of birth weights, cf. Abreveya (2001). Of special interest, however, are the very low quantiles, since low birth weights have been linked to subsequent health problems. This paper provides inference tools for statistical investigation of this question using quantile regressions.

Example 3. Probabilistic Frontiers for Producers. An important form of efficiency analysis in industrial organization and economics of regulation is the determination of production frontiers (e.g. Timmer, 1971). Given production factors W, high percentiles of production levels Y, called probabilistic frontiers, namely $Q_Y(\tau|W)$, for $\tau \in [1-\epsilon, 1]$ and $\epsilon > 0$ are attained by an ϵ-fraction of most highly efficient firms. The inference on ϵ-frontiers (**??**) will be facilitated by the results of this paper.

Example 4. (S,s)-Rules and Other Approximate Reservation Rules in Economic Decisions. A related example is that of (S, s)-adjustment models, which arise as optimal policies in many economic models (Arrow, Harris, and Marschak 1951). For example, the capital Z is adjusted up to the level S once it has depreciated to the level s. In terms of an econometric specification, the observed capital satisfies $Z_i = s(W_i) + v_i$, where W_i are covariates and v_i is a disturbance that is positive most of the time, i.e. $P(v_i \geq 0)$ is close to 1. Once capital reaches the critical level, $Z_i - s(W_i) \leq 0$, it is adjusted in the next period. When the disturbance v_i is negative, it is assumed to capture unobserved heterogeneity and small decision mistakes which are independent of observed covariates W_i, as in Caballero and Engel (1999). In a given cross-section or time series, adjustment will occur infrequently, so in fact data at or below the lower adjustment band $s(X_i)$ will be observed with a small probability $P(v_i \leq 0)$, hence $Q_Z(\tau|W) = s(W) + Q_v(\tau)$ for $\tau \in (0, P(v \leq 0)]$. The lower band function $s(W)$ therefore coincides with the lower conditional quantile function up to an additive constant. A similar argument can be made for the upper band function $S(W)$.

Example 5. Structural Auction Models. In the standard specification of the first-price procurement auction where bidders hold independent valuations, the winning bids B_i satisfy: $B_i = c(X_i)\beta(n_i) + \epsilon_i, \epsilon_i \geq 0$, where $c(X_i)$ is the efficient cost function, and $\beta(n_i) \geq 1$ is a mark-up that approaches 1 as the number of bidders n_i approaches infinity (Donald and Paarsch 2002). By construction, $c(X_i)\beta(n_i)$ is the extreme conditional quantile function. In the empirical analysis, it is realistic to let the disturbance ϵ_i take on some small negative value, so that when negative these disturbances capture small decision mistakes that are independent of included explanatory variables. In this case the quantile function satisfies $Q_B(\tau|X, n) = C(X)\beta(n) + Q_\epsilon(\tau)$, for $\tau \in (0, P[\epsilon \leq 0]]$, hence inference on $c(X)$ and $\beta(n)$ can be made using the extremal QR, complementing the approach of Donald and Paarsch (2002).[1]

2.2. **The Econometric Problem.** The linear-in-parameters QR model specifies a τ-th quantile of interest of a response variable Y given regressors X as $X'\beta(\tau)$. Suppose we have $T = 200$ observations and run QR, obtain $\widehat{\beta}(\tau)$, and proceed with inference on $\beta(\tau)$ for $\tau = 5\%$. For inference purposes, we may use the conventional central normal limit theory which states

$$\sqrt{T}(\widehat{\beta}(\tau) - \beta(\tau)) \to_d N\left(0, \Omega(\tau)\right), \text{ for fixed } \tau \in (0,1), \text{ as } T \to \infty, \tag{2.1}$$

for $\Omega(\tau) = \tau(1-\tau)J(\tau)^{-1}E[XX']J(\tau)^{-1}, J(\tau) = E[f_U(F_U^{-1}(\tau|X)|X)XX']$, and $U = Y - X'\beta$ for some $\beta \in \mathbb{R}^d$.

In this paper we develop an inference method based on a limit theory obtained under the *extreme order conditions* on the quantile index – $\tau T \to k > 0$ as $T \to \infty$. The idea is that this fits our example with $\tau T = 10$ and other similar cases better than the normal approach arising from imposing $\tau T \approx \infty$. Indeed, the canonically normalized QR (CN-QR) satisfies:

$$A_T(\widehat{\beta}(\tau) - \beta(\tau)) \to_d \widehat{Z}_\infty(k), \quad \text{for } \; A_T = \frac{1}{F_U^{-1}(1/T)}, \text{ as } \; \tau T \to k, \; T \to \infty, \tag{2.2}$$

where $\widehat{Z}_\infty(k)$ is an extreme-type variate specified in Section 3 that depends on the tail thickness parameter ξ of variable U. The normalization by A_T is canonical in the sense

---

[1] In the approach of Donald and Paarsch (2002) one first estimates a saturated boundary model using cell-by-cell extreme order statistics and then projects the estimated boundary onto a lower dimensional non-linear model using the minimum-distance method. One can proceed likewise using quantile regression to estimate the first stage and then project it onto further structural model.

that the same normalization is used in the classical univariate theory of extreme order statistics.

Figure 1 shows that the extremal approximation (2.2) may *potentially* provide an improved basis for inference in the tails. It explains the finite sample distribution of $\widehat{\beta}(\tau)$ in the tails considerably better than the normal distribution does. However, a major obstacle to this route and to adaptation of this approximation in practice is the feasibility: we need to have accurate estimates of the tail index parameter ξ and of the normalization factor $1/F_U^{-1}(1/n)$. The scaling $1/F_U^{-1}(1/T)$ is a reciprocal of an extreme population quantile which is hard or infeasible to estimate consistently in the heavy-tailed cases. Under a minimal assumption that disturbances belong to the domain of minimum attraction, we have that $1/F_U^{-1}(1/T) = L(1/T) \cdot T^{-\xi}(1 + o(1))$ where $L(\cdot)$ is a nonparametric, slowly varying function, which is too weak a restriction to be useful (Bertail, Haefke, Politis, and White 2004). Estimation of $L(1/T)$ becomes feasible when parametric assumptions on $L(\cdot)$ are made; and some will be given later. Estimation of tail parameters such as the extreme value index ξ is another known difficult problem, as the common estimators behave well only when the slowly-varying component $L(1/T)$ is not present ((Embrechts, Klüppelberg, and Mikosch 1997)).

As neither the normalization $1/F_U^{-1}(1/T)$ nor the tail parameters ζ are of primary interest, it is reasonable to attempt to bypass the estimation of these quantities. In order to do so, our inference will be based on the Self-Normalized QR statistic (SN-QR)

$$\mathcal{A}_T \left(\widehat{\beta}(\tau) - \beta(\tau)\right) \to_d Z_\infty(k), \text{ for } \mathcal{A}_T := \frac{\sqrt{\tau T}}{\bar{X}'(\widehat{\beta}(m\tau) - \widehat{\beta}(\tau))}, \tag{2.3}$$

where $Z_\infty(k)$ is another extreme type variate, to be specified later. The random scaling $\mathcal{A}_T$ is function of data only, and it is not an estimate of A_T. Using this result we will provide two inference approaches: an analytical approach and a resampling approach based on modified subsampling. The analytical approach will directly rely on the (simulated) limit distribution, but will require estimation of the tail parameter ξ. However, our resampling approach eliminates the need to estimate ξ.

Our subsampling approach is different from conventional subsampling in its use of recentering terms and of random normalization. The 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 $A_T \to 0$ when $\xi > 0$. Instead, we will employ recentering by intermediate order QR estimates, which will be consistent enough to estimate the limit distribution of SN-QR consistently. Thus, the modified 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.

3. The Set Up and the Model

3.1. **Basic Set Up.** Suppose Y is the response variable in $\mathbb{R}$, and W is a vector of the conditioning variables. Denote the conditional distribution of Y given $W = w$, $F_Y(\cdot|w)$. The τ-th conditional quantile function $F_Y^{-1}(\tau|w)$ is the inverse of $F_Y(\cdot|w)$ evaluated at probability index τ. We use the following flexible functional form for the extremal conditional quantile function of Y given W:

$$F_Y^{-1}(\tau|X) := X'\beta(\tau), \quad (\forall \tau \in \mathcal{I} = (0,\eta), \exists \eta > 0), \tag{3.4}$$

where $X = B(W)$ is a $d \times 1$ vector of approximating functions that may include power functions, splines, and other transformations of original variables W. As noted in Hedricks and Koenker (1992) and Newey (1997), economic theory rarely prescribes exact functional forms motivating the use of flexible functional forms to provide parsimonious approximations of unknown economic structures and the computational convenience.

Given T observations $\{Y_t, X_t, t = 1, ..., T\}$, the quantile regression (QR) solves:

$$\widehat{\beta}(\tau) \in \arg\min_{\beta \in \mathbb{R}^d} \sum_{t=1}^{T} \rho_\tau\left(Y_t - X_t'\beta\right), \tag{3.5}$$

where $\rho_\tau(u) = (\tau - 1(u < 0))u$. The median $\tau = 1/2$ case of (3.5) was introduced by Laplace (1818) and the general quantile formulation (3.5) by Koenker and Bassett (1978).

Viewing the sample regression quantiles as order statistics in the regression setting, we will refer to τT as the *order* or *rank* of QR. The sequence of quantile index-sample size pairs (τ_T, T) will be said to be an extreme order sequence, if $\tau_T \searrow 0$ and $\tau_T T \to k > 0$, an intermediate order sequence, if $\tau_T \searrow 0$ and $\tau_T T \to \infty$, and a central order sequence, if τ is fixed and $T \to \infty$. Our inference analysis is based on extreme-order sequences that provide non-normal approximations to the distribution of QR statistics.

3.2. **The Extremal Quantile Regression Model.** In developing our inference methods, we assume that the response variable Y has Pareto-type tails. The Pareto-type tails are regularly varying tails, and are prevalent in economic data, as discovered by Vilfredo Pareto in 1895.[2] The 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.

---

[2] Pareto called the tails of this form “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 Gnedenko and de Haan.

Consider a random variable u with the distribution function F_u and the lower end-point $s = -\infty$ or $s = 0$. F_u is said to be *regularly varying* at s, if

$$\lim_{l \searrow s} \frac{F_u(lv)}{F_u(l)} = v^{-1/\xi}, \ (\forall v > 0), (\exists \ \xi > 0 \text{ or } \xi < 0). \tag{3.6}$$

Regular variation equivalently means that the distribution function F_u and its quantile function F_u^{-1} exhibit approximately power-law behavior in the tails:[3]

$$\begin{aligned} F_u(t) &\sim \bar{L}(l) \cdot l^{-1/\xi}, \text{ as } l \searrow s, & (3.7) \\ F_u^{-1}(\tau) &\sim L(\tau) \cdot \tau^{-\xi}, \text{ as } \tau \searrow 0, & (3.8) \end{aligned}$$

where $\bar{L}(l)$ is a nonparametric, slowly-varying function at s and $L(\tau)$ is a nonparametric slowly varying function at 0. A function $l \mapsto L(l)$ is slowly varying at s if $\lim_{l \searrow s}[L(l)/L(ml)] = 1$ for any $m > 0$, for example, $L(l) = L$, $L(l) = L \cdot \log(-l)$, etc.

The number ξ is called the extreme value (EV) index and its absolute value $|\xi|$ measures heavy-tailedness of distributions. A distribution F_u with regularly varying tails necessarily has a finite lower support point if $\xi < 0$ and infinite lower support point if $\xi > 0$. Distributions with $\xi > 0$ include stable distributions, Pareto distributions, t distributions, and others. For instance, the t-distribution with ν degrees of freedom has the EV index $\xi = 1/\nu$, where $\xi = 1$ yields the Cauchy distribution that has heavy tails, and $\xi < 1/30$ closely approximates the normal distribution that has light tails. On the other hand, distributions with $\xi < 0$ include the uniform, exponential, and Weibull distributions, among others.

Our main assumption is that the response variable Y, transformed by some auxiliary regression line, has regularly varying tails at $s = -\infty$ and 0 with EV index ξ.

C1. *In addition to (3.4), suppose there exists an auxiliary extremal regression parameter* β_e *such that* $U \equiv Y - X'\beta_e$ *has end-point* $s = 0$ *or* $s = -\infty$ *a.s. and satisfies as* $\tau \searrow 0$

$$F_U^{-1}(\tau|x) \sim x'\mathbf{c} \cdot F_u^{-1}(\tau), \ \textit{uniformly in } x \in \mathbf{X}, \ \textit{where } F_u^{-1}(\tau) \in RV_{-\xi}(0). \tag{3.9}$$

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 in the tail is approximately given by $\beta_e + \mathbf{c}F_u^{-1}(\tau)$, which allows for a differential impact of covariates across various extremal quantiles.

C2. *In addition to C1,* $\partial F_U^{-1}(\tau|x)/\partial\tau$ *exists and* $\partial F_U^{-1}(\tau|x)/\partial\tau \sim x'\mathbf{c} \cdot \partial F_u^{-1}(\tau)/\partial\tau$ *as* $\tau \searrow 0$*, uniformly in* $x \in \mathbf{X}$*, where* $F_u^{-1}(\tau)/\partial\tau \in RV_{-\xi-1}$ *(0).*

This assumption strengthens C1 by imposing regular variation on the quantile density function, which is important for obtaining the main inferential results.

---

[3] The notation $a \sim b$ means that $a/b \to 1$ as appropriate limits are taken.

3.3. **Sampling Conditions.**

The following sampling conditions will be imposed.

C3. *The regressor X is of the form $(1, Z')'$, has compact support $\mathbf{X}$, mean $\mu_X = (1, 0, ...)'$ and $EXX' > 0$, and satisfies the non-lattice condition stated in the mathematical appendix (satisfied, for instance, if Z is absolutely continuous).*

Compactness is a necessary technical assumption needed to insure 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 only that data for which X belongs to a compact region. The non-degeneracy condition is a standard condition that requires invertibility of EXX'. The non-lattice condition is required for existence of finite-sample density of QR coefficients and is needed here even asymptotically, since the asymptotic distribution theory closely resembles the finite-sample theory, which is not a surprise given the rare nature of events that have the probability of order $1/T$.

Data is also assumed to satisfy the canonical iid sampling.

C4. *The sequence of variables $\{(Y_t, X_t), t \geq 1\}$ are independent and identically distributed.*

More generally, data can be a time series with the following restriction.

C5. *The sequence of vectors $\{(Y_t, X_t), t \geq 1\}$ forms a stationary, α-mixing process with geometric mixing rate and the extreme events satisfying a non-clustering condition: $P_t(U_t \leq K, U_{t+j} \leq K) \leq CP_t(U_t \leq K)^2$ for all sufficiently small $K \geq s$, all $j \geq 1$, all $t \geq 1$, and some constant $C > 0$, where P_t denotes the probability conditional on the information set up to time t (that is, the σ-algebra generated by $\{Y_{t'}, X_{t'}, t' = t-1, t-2, ...\}$).*

The assumption of mixing is a standard assumption made in econometric analysis, e.g. White. The non-clustering condition of the Meyer (1973)-type states that the possibility of two extreme events co-occuring at fixed dates is much lower than 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 the 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 robust to a violation of this assumption.

4. Asymptotic Extreme Value Approximations: Feasible and Infeasible

This section establishes important preliminary results that will underlie our inferential procedures.

4.1. **Weak Limits of Extreme Order QR and Self-Normalized QR statistics.** The asymptotic distributions are given under the extreme order sequence $\tau T \to k > 0$ as $T \to 0$. Define the self-normalized QR (SN-QR) statistic as:

$$Z_T(k) := \mathcal{A}_T\Big(\widehat{\beta}(\tau) - \beta(\tau)\Big) \text{ for } \mathcal{A}_T := \frac{\sqrt{\tau T}}{\bar{X}'(\widehat{\beta}(m\tau) - \widehat{\beta}(\tau))}, \tag{4.10}$$

where $\tau T(m-1) > d$, and the canonically-normalized QR statistic (CN-QR) as:

$$\widehat{Z}_T(k) := A_T\Big(\widehat{\beta}(\tau) - \beta(\tau)\Big) \text{ for } A_T := 1/F_u^{-1}\left(\frac{1}{T}\right). \tag{4.11}$$

The first statistic uses a random normalization, while the second uses an infeasible canonical normalization.

Theorem 1 (Weak Limits). *Suppose conditions C1-C5 hold. Define*

$$\widehat{Z}_\infty(k) \equiv \arg\min_{z \in \mathbb{R}^d} \left[-k\mu_X'(z - k^{-\xi}\mathbf{c}) + \sum_{i=1}^{\infty} [\Gamma_i^{-\xi} \cdot \mathcal{X}_i'\mathbf{c} + \mathcal{X}_i'(z - k^{-\xi}\mathbf{c})]_+ \right], \tag{4.12}$$

where $\{\Gamma_1, \Gamma_2, ...\} \equiv \{\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, ...\}$, $\{\mathcal{E}_1, \mathcal{E}_2, ...\}$ *is an iid sequence of exponential variables, which is independent of* $\{\mathcal{X}_1, \mathcal{X}_2, ...\}$, *an iid sequence with distribution* F_X. *Then,*

1. *The SN-QR statistics of order* k *behave as follows, for any* m *such that* $k(m-1) > d$*: as* $\tau T \to k > 0$ *and* $T \to \infty$*, we have that* $Z_T(k) \to_d Z_\infty(k)$*, where*

$$Z_\infty(k) := \frac{\sqrt{k}\widehat{Z}_\infty(k)}{\mu_X'(\widehat{Z}_\infty(mk) - \widehat{Z}_\infty(k)) + (m^{-\xi} - 1)k^{-\xi}}. \tag{4.13}$$

2. *The CN-QR statistics of order* k *behave as follows: as* $\tau T \to k > 0$ *and* $T \to \infty$*,* $\widehat{Z}_T(k) \to_d \widehat{Z}_\infty(k)$*.*

Comment 4.1. The condition that $k(m-1) > d$ insures that $\mathcal{A}_T$ is well defined, thanks to Theorem 3.2 in Bassett and Koenker (1982). Result 1 on SN-QR statistics is the main new result that we will need for inference. Result 2 on CN-QR statistics is needed primarily for auxiliary purposes. Chernozhukov (2000b) presents some extensions of Result 2. Previously, Feigin and Resnick (1994), Portnoy and Jurečková (1999), and Knight (1999) investigated canonically normalized statistics under the limit case $\tau T \to 0^+$ (known as the linear programming estimator), under various degrees of generality. The near-extreme case $\tau T \to k > 0$ considered here is motivated primarily by economic applications listed in Section 2.

4.2. **Discussion.** Few basic observations are in order. Recall that in the classical non-regression case, when $X = 1$, and U_t are sampled iid according to F_u, we have that $A_T(\widehat{\beta}(\frac{k}{T}) - \beta(\frac{k}{T}))$, $k = 1, ..., J$, converge in distribution to $\Gamma_k^{-\xi} - k^{-\xi}$, $k = 1, ..., J$, where $\widehat{\beta}(\frac{k}{T}) := U_{(k)}$ is the k-th order statistic and the $\lceil k/T \rceil$-th sample quantile. These basic variables also appear in the limiting expressions for SN-QR and CN-QR statistics with a prefactor $\mathcal{X}_i'\mathbf{c}$ that makes them correlated with regressors $\mathcal{X}_i$. The limit distributions are not normal; they have no finite moments of any order exceeding $1/\xi$ when $\xi > 0$; they are not centered at zero, and significant first order asymptotic median biases may exist. It is therefore desirable to construct statistics that are asymptotically median-unbiased.

In contrast to limit theory for SN-QR statistics, Result 1, the limit theory of CN-QR statistics, Result 2, is generally infeasible for purposes of inference, since it requires knowledge or estimability of the scaling constant $A_T = 1/F_u^{-1}(\frac{1}{T})$, the reciprocal of the extremal quantile of variable U defined in C1. That is, one requires $\widehat{A}_T$ such that $\widehat{A}_T/A_T \to_p 1$, which is not generally feasible unless further strong parametric restrictions on the tail are made. Even though these restrictions facilitate estimation of A_T and hence inference on CN-QR, as described later, this will not be our primary inferential method.

The last point is important. By assumption, $A_T = 1/F_u^{-1}\left(\frac{1}{T}\right) = L(\frac{1}{T})T^{-\xi}$, where $L(\frac{1}{T})$ is a nonparametric slowly-varying function. The class of slowly varying functions is quite broad: by Karamata's Representation $L(\tau) = \ell(\tau)\exp(\int_1^{1/\tau} t^{-1}\varepsilon(t)dt)$, where $\lim_{t\searrow 0}\ell(t) = \ell \in (0, 1)$, $\lim_{t\searrow 0}\varepsilon(t) = 0$, and $\ell(t)$ and $\varepsilon(t)$ are measurable, nonnegative functions, cf. Resnick (1987). Nonparametric estimation of $L(\tau)$ at $\tau = 1/T$ under these restrictions and using the data sample of size T appears to be an infeasible problem (e.g. Bertail, Haefke, Politis, and White (2004)).

Our main proposal for inference is based on SN-QR and Result 1, and it entirely avoids estimating A_T. We will estimate the distribution of SN-QR statistics $\mathcal{A}_T(\widehat{\beta}(\tau) - \beta(\tau))$ instead, using either a subsampling method or an analytical method. An important ingredient here is the feasible normalizing variable $\mathcal{A}_T$ that 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.[4] An advantage of the subsampling method over analytical methods is that does not require estimation of nuisance parameters ξ and $\mathbf{c}$.

---

[4] The idea of feasible random normalization has been used in many other contexts (e.g. t-statistics). In extreme value theory, Dekkers and de Haan (1989) applied a similar random normalization idea to the extrapolated quantile estimators in the non-regression setting, precisely to produce limit distributions that can be easily used for inference.

4.3. **Generic Inference and Median-Unbiased Estimation.** Next we outline procedures for conducting inference and constructing asymptotically median unbiased estimates of linear functions $\psi'\beta(\tau)$ of the coefficient vector $\beta(\tau)$ for extremal values of τ.

Inference Using SN-QR. By Theorem 1 for any non-zero ψ, $\psi'\mathcal{A}_T(\widehat{\beta}(\tau) - \beta(\tau)) \to_d \psi' Z_\infty(k)$. Given a consistent estimate $\widehat{c}_\alpha$ of the α-quantile c_α of $\psi' Z_\infty(k)$, the asymptotically median-unbiased estimator and the α%-confidence interval can be constructed as

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

where the bias-correction term $c_{1/2}/\mathcal{A}_T$ and components $c_{\alpha/2}/\mathcal{A}_T$ and $c_{1-\alpha/2}/\mathcal{A}_T$ of the confidence interval depend on the random scaling $\mathcal{A}_T$. Consistent estimates $\widehat{c}_\alpha$ are provided in the next section.

Theorem 2 (Inference and median-unbiased estimation using SN-QR)**.** *Under conditions of Theorem 1, suppose we have* $\widehat{c}_\alpha$ *such that* $\widehat{c}_\alpha \to_p c_\alpha$, *where* c_α *is the* α*-quantile of* $\psi' Z_\infty(k)$. *Then,* $\lim_{T\to\infty} P\{\psi'\widehat{\beta}(\tau) - \widehat{c}_{1/2}/\mathcal{A}_T \leq \psi'\beta(\tau)\} = 1/2$ *and* $\lim_{T\to\infty} P\{\psi'\widehat{\beta}(\tau) - \widehat{c}_{\alpha/2}/\mathcal{A}_T \leq \psi'\beta(\tau) \leq \psi'\widehat{\beta}(\tau) - \widehat{c}_{1-\alpha/2}/\mathcal{A}_T\} = \alpha$.

Inference Using CN-QR. Under the conditions stated above, for any non-zero ψ, $\psi' A_T(\widehat{\beta}(\tau) - \beta(\tau)) \to_d \psi'\widehat{Z}_\infty(k)$. Given a consistent estimate $\widehat{A}_T$ and a consistent estimate $\widehat{c}'_\alpha$ of the α-quantile c'_α of $\psi'\widehat{Z}_\infty(k)$, we could construct the asymptotically median-unbiased estimator and α%-confidence intervals as, respectively:

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

Construction of consistent estimates of A_T requires additional strong, nearly parametric restrictions on the underlying model. For example, suppose that the nonparametric slowly varying component $L(\tau)$ of A_T is replaced by a constant L, i.e.

$$1/F_u^{-1}(\tau) = L \cdot \tau^\xi \cdot (1 + \delta(\tau)), \ \delta(\tau) = o(1), \ L \in \mathbb{R}, \text{ as } \tau \searrow 0. \tag{4.14}$$

Then, one can estimate the constants L and ξ via Pickands-type procedures:

$$\widehat{\xi} = \frac{-1}{\ln 2} \ln \frac{\bar{X}'(\widehat{\beta}(4\tau_T) - \widehat{\beta}(\tau_T))}{\bar{X}'(\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T))} \text{ and } \widehat{L} = \frac{\bar{X}'(\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T))}{(2^{-\widehat{\xi}} - 1) \cdot T^{-\widehat{\xi}}}, \tag{4.15}$$

where τ_T is chosen to of an intermediate order, $\tau_T T \to \infty$ and $\tau_T \to 0$. By Theorem 4 in Chernozhukov (2000b), these estimates are consistent under C1-C5, and under further conditions on the sequence $(\delta(\tau_T), \tau_T)$,[5] one has that $\widehat{\xi} = \xi + o(1/\ln T)$, which produces a consistent estimate $\widehat{A}_T = \widehat{L}(1/T)^{-\widehat{\xi}}$ such that $\widehat{A}_T/A_T \to_p 1$.

Consistent estimates of c'_α are provided in the next section.

Theorem 3 (Inference and Median-Unbiased Estimation using CN-QR)**.** *Assume conditions of Theorem 1 hold. Suppose that we have* $\widehat{A}_T$ *such that* $\widehat{A}_T/A_T \to_p 1$ *and* $\widehat{c}'_\alpha$ *such that* $\widehat{c}_\alpha \to_p c'_\alpha$, *where* $c'_{\alpha'}$ *is the* α*-quantile of* $\psi'\widehat{Z}_\infty(k)$. *Then,* $\lim_{T\to\infty} P\{\psi'\widehat{\beta}(\tau) - \widehat{c}'_{1/2}/\widehat{A}_T \leq \psi'\beta(\tau)\} = 1/2$ *and* $\lim_{T\to\infty} P\{\psi'\widehat{\beta}(\tau) - \widehat{c}'_{\alpha/2}/\widehat{A}_T \leq \psi'\beta(\tau) \leq \psi'\widehat{\beta}(\tau) - \widehat{c}_{1-\alpha/2}/\widehat{A}_T\} = \alpha$.

5. Estimation of Critical Values

5.1. **Subsampling-Based Estimation of Critical Values.** The method developed below uses subsamples to estimate the sampling distribution of SN-QR, just as in standard subsampling. However, by using SN-QR, the method bypasses estimation of the unknown convergence rate A_T, required in standard subsampling. The method also employs recentering terms that avoid the inconsistency of the standard subsampling caused by divergence of the QR estimator when $\xi > 0$.

The method is as follows. First, consider all subsets of data $\{W_t = (Y_t, X_t),\ t = 1, ..., T\}$ of size b. If $\{W_t\}$ is a time series, consider $B_T = T - b + 1$ subsets of size b of the form $\{W_i, ..., W_{i+b-1}\}$. Then compute the analogs of SN-QR statistics in subsamples, denoted $\widehat{V}_{i,b}$, and defined below in equation (5.17), for each i-th subsample, for $i = 1, ..., B_T$. Second, estimate c_α by $\widehat{c}_\alpha$ defined as the sample α-quantile of $\{\widehat{V}_{i,b,T}, i = 1, ..., B_T\}$. In practice, a smaller number B_T of randomly chosen subsets can be used, provided $B_T \to \infty$ as $T \to \infty$, see Sec. 2.5 in Politis et. al. (1999). Subsample size b can be chosen according to Politis et. al. (1999) and Bertail et. al. (2004).

Recall that the SN-QR statistic computed in the full sample of size T is:

$$V_T \quad := \mathcal{A}_T\psi'(\widehat{\beta}_T(\tau_T) - \beta(\tau_T)),\ \ \mathcal{A}_T := \sqrt{\tau_T T}/\bar{X}'\big(\widehat{\beta}(m\tau_T) - \widehat{\beta}(\tau_T)\big), \tag{5.16}$$

where m will be set equal $1.5 - 3$.[6] In this section we write τ_T to emphasize the theoretical dependence of the quantile of interest τ on the sample size. In each i-th subsample of size b, we compute the following analog of the above statistic:

$$\widehat{V}_{i,b,T} \quad := \mathcal{A}_{i,b,T}\psi'(\widehat{\beta}_{i,b,T}(\tau_b) - \widehat{\beta}(\tau_b)),\ \ \mathcal{A}_{i,b,T} := \sqrt{\tau_b b}/\bar{X}'\big(\widehat{\beta}_{i,b,T}(m\tau_b) - \widehat{\beta}_{i,b,T}(\tau_b)\big), \tag{5.17}$$

where $\widehat{\beta}(\tau_T)$ is the τ_T-regression quantile coefficient computed using the full sample; $\widehat{\beta}(\tau_b)$ is the τ_b-regression quantile coefficient computed using the full sample; $\widehat{\beta}_{i,b,T}(\tau_b)$ is the

---

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

[6] Such choice, namely $m = 2$, is often used in the non-regression cases, cf. Dekkers and de Haan (1989).

τ_b-regression quantile coefficient computed using i-th subsample; $\widehat{\beta}_{i,b,T}(m\tau_b)$ is the $m\tau_b$-regression quantile coefficient computed using i-th subsample; and $\tau_b := k_T/b := (\tau_T T)/b$.[7] Note that determination of τ_b is a critical decision that sets apart the central order approximation from the extreme order approximation. In the former case case one sets $\tau_b = \tau_T$ in subsamples.

Under additional parametric assumptions on the tail behavior stated earlier, one can also estimate the quantiles of the limit dsitribution of CN-QR. The resampling procedure can be done in two steps: First, create subsamples $i = 1, ..., B_T$ as before, compute in each subsample: $V_{i,b,T} := \widehat{A}_b \psi'(\widehat{\beta}_{i,b,T}(\tau_b) - \widehat{\beta}(\tau_b))$, where $\widehat{A}_b$ is any consistent estimate of A_b. Estimate c_α by $\widehat{c}_\alpha$ defined as the α-quantile of $\{V_{i,b,T}, i = 1, ..., B_T\}$. For example, under the parametric restrictions specified in (4.14), set $\widehat{A}_b = \widehat{L} b^{-\widehat{\xi}}$ for $\widehat{L}$ and $\widehat{\xi}$ specified in (4.15).

The following theorems establish consistency of $\widehat{c}_\alpha$ and $\widehat{c}'_\alpha$.

Theorem 4 (Critical Values for SN-QR by Resampling)**.** *Given the assumptions of Theorems 1 and 2 and that $b/n \to 0, b \to \infty, n \to \infty, B_n \to \infty$, we have that $\widehat{c}_\alpha \to_p c_\alpha$.*

Theorem 5 (Critical Values for CN-QR by Resampling)**.** *Suppose that the assumptions of Theorems 1 and 2 hold and $b/n \to 0, b \to \infty, n \to \infty, B_n \to \infty$, and the estimate $\widehat{A}_b$ is such that $\widehat{A}_b/A_b \to 1$. Then $\widehat{c}'_\alpha \to_p c'_\alpha$.*

Comment 5.1. The method produces consistent inference in the QR case and might be of independent interest in the non-regression case. The method differs from conventional subsampling in several ways: First, our method also uses random normalization term $\mathcal{A}_T$. Second, QR statistics are diverging when $\xi > 0$. In this case, the standard subsampling that employs the recentering by the full sample estimate $\widehat{\beta}(\tau_T)$ does not apply. In order to use this re-centering one requires $A_b/A_T \to 0$ (Politis and Romano 1994), but instead $A_b/A_T \to \infty$. Our approach instead uses $\widehat{\beta}(\tau_b)$ for recentering. This statistic itself may be diverging, but the speed of its divergence is strictly slower than A_T, because it is an intermediate order QR statistic. Hence this method of recentering exploits the special structure of order statistics.

5.2. **Analytical Estimation of Critical Values.**

Analytical inference uses the quantiles of the limit distributions found in Theorem 1, which can only be done numerically via simulation procedures. This approach is much more demanding than the previous subsampling

---

[7] In practice, it is reasonable to use the following finite-sample adjustment to τ_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 if $\tau_T > .2$, it is judged to be non-extremal, and the subsampling procedure resorts to the central order inference. If $\tau_T < .2$, then $\tau_b = \min[(\tau_T T)/b, .2]$. The truncation of τ_b by .2 is a finite-sample adjustment that restricts the key statistics $\widehat{V}_{i,b,T}$ to be extremal in subsamples. Adjustments of this kind do not affect the asymptotic arguments.

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

Define the following random vector

$$\widehat{Z}^*_\infty(k) = \arg\min_{z\in\mathbb{R}^d}[-k\widehat{\mu}'_X(z - k^{-\widehat{\xi}}\widehat{\mathbf{c}}) + \sum_{i=1}^{\infty}(\Gamma_i^{-\widehat{\xi}}\cdot\mathcal{X}_i'\widehat{\mathbf{c}} + \mathcal{X}_i'(z - k^{-\widehat{\xi}}\widehat{\mathbf{c}}))_+] \tag{5.18}$$

for some estimates $\widehat{\xi}$ and $\widehat{\mathbf{c}}$, for example those stated below, $\widehat{\mu}_X = \bar{X}$, $\{\Gamma_1, \Gamma_2, ...\} = \{\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, ...\}$, where $\{\mathcal{E}_1, \mathcal{E}_2, ...\}$ is an i.i.d. sequence of unit-exponential variables, and $\{\mathcal{X}_1, \mathcal{X}_2, ...\}$ is an iid sequence with distribution function $\widehat{F}_X$, where $\widehat{F}_X$ is a "smoothed" consistent estimate of F_X. Also, let $Z^*_\infty(k) = \sqrt{k}\widehat{Z}^*_\infty(k)/[\bar{X}'(\widehat{Z}^*_\infty(mk) - \widehat{Z}^*_\infty(k)) + |m^{-\widehat{\xi}} - 1|k^{-\widehat{\xi}}]$. Estimates $\widehat{c}'_\alpha$ and $\widehat{c}_\alpha$ are obtained by taking α-quantiles of variables $\psi'\widehat{Z}^*_\infty(k)$ and $\psi' Z^*_\infty(k)$. In practice, these quantiles can only be evaluated numerically as described below.

The analytical inference procedure requires consistent estimates of ξ and $\mathbf{c}$. The consistent estimates based on Pickands type procedures are provided in Chernozhukov ((2000b), Theorem 5) under conditions implied by Assumption 1:

$$\widehat{\xi} = \frac{-1}{\ln 2}\ln\frac{\bar{X}'(\widehat{\beta}(4\tau_T) - \widehat{\beta}(\tau_T))}{\bar{X}'(\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T))} \text{ and } \widehat{\mathbf{c}} = \frac{\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T)}{\bar{X}'(\widehat{\beta}(2\tau_T) - \widehat{\beta}(\tau_T))}, \tag{5.19}$$

where $\tau_T T \to \infty$ and $\tau_T \to 0$. Also, the empirical distribution function $\widehat{F}_X$ is a uniformly consistent estimate of F_X via the Glivenko-Cantelli Theorem.

Theorem 6 (Critical Values for CN-QR by Analytical Method)**.** *Assume conditions of Theorem 1 hold. Then, for any consistent estimates of nuisance parameters such that* $\widehat{\xi} \to_p \xi$, $\widehat{\mathbf{c}} \to_p \mathbf{c}$, *such as the ones provided above,* $\widehat{c}'_\alpha = c'_\alpha + o_p(1)$.

Theorem 7 (Critical Values for SN-QR by Analytical Method)**.** *Assume conditions of Theorem 1. Then for any consistent estimates of nuisance parameters such that* $\widehat{\xi} \to_p \xi$, $\widehat{\mathbf{c}} \to_p \mathbf{c}$, *such as the ones provided above,* $\widehat{c}_\alpha = c_\alpha + o_p(1)$.

Comment 5.2. Since the distributions of the variables $\widehat{Z}_\infty(k)$ and $Z_\infty(k)$ are not in closed form, except in very special cases, we can only obtain $\widehat{c}'_\alpha$ and $\widehat{c}_\alpha$ numerically via a Monte Carlo procedure. For instance, for each $i = 1, ..., B$ the procedure $\widehat{Z}^*_{i,\infty}(k)$ using the formula ( ) above except truncating the sum at some value L. We find that choosing $L \geq 200$ and setting $B \geq 100$ provides numerically accurate estimates (see Chernozhukov (2000a)).

6. Extreme Value vs. Normal Inference: Comparisons

6.1. **Properties of Confidence Intervals with Unknown Nuisance Parameters.** In this section we asses the covering performance of normal and extremal CI using the model

$Y_t = X_t'\beta + U_t$, $t = 1, ..., 500$, $\beta = 1$, where 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 $\alpha \in \{.5, 1, 4\}$. These distributions have the extreme value index $\xi = 1/\nu \in \{1, 1/3, 1/20\}$ and $\xi = -1/\alpha \in \{-2, -1, -1/4\}$, respectively. Regressors are drawn with replacement from the empirical application in Section 7.1 in order to match a real situation as closely as possible.[8] 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 CI based on our primary extremal method and on the normal inference method suggested by Powell (1986). The figures are based on QR estimates for $\tau \in \{0.005, .02, .05, .1\}$, i.e. $\tau T \in \{2.5, 10, 25, 50\}$.

When disturbances U are t-distributed, extremal CI have good coverage properties, whereas normal CI typically *undercover* with the performance that deteriorates in the degree of heavy-tailedness and improves in the index τT. In heavy-tailed cases, $\xi \in \{1, 1/3\}$, normal CI substantially undercover, as might be expected from the normal distribution failing to capture heavy tails of the actual distribution of QR statistics. In thin-tailed cases, $\xi = 1/30$, normal CI have an improved coverage of $70 - 80\%$. In all cases extremal CI do better consistently, giving coverage in a good neighborhood of 90%.

When disturbances U are Weibull-distributed, extremal CI also have good coverage properties, whereas normal CI typically *overcover* with the performance that deteriorates in the degree of heavy-tailedness and improves in the index τT. In heavy-tailed cases, $\xi \in \{-2, -1\}$, normal CI strongly overcover, which results from the over-dispersion of the normal distribution relative to the actual distribution of QR statistics. In thin-tailed cases, $\alpha = 4$, normal CI cover at the right rate, even though formally they are not valid in the tails. In all cases, extremal CI do better or as well as normal CI, giving a coverage in a good neigborhood of 90%.

We have also compared forecasting properties of ordinary QR estimators and median-bias-corrected QR estimators of the true intercept and slope coefficients, using the median absolute deviation and median bias as measures of the performance (other measures may be infinite). We have found that gains to bias-correcting appear to be very small, except in the finite-support case with disturbances heavy-tailed near the boundary. Tabulations of these results are therefore omitted.

6.2. **A"Rule-of-Thumb" for Application of Extremal Inference.** Equipped with both simulation experiments and practical experiences, we would like to find a simple rule for the application of the extremal inference. To this end, note that the order τT of the QR

---

[8] These data can be downloaded at www.mit.edu/vchern.

statistic $\widehat{\beta}(\tau)$ should obviously play the crucial rule, as one needs $\tau T \to \infty$ to have normal laws apply. The number of regressors may also play a crucial role. Consider the case where all d regressors, other than the intercept, are indicators that equally divide the sample of size T into subsamples of size T/d. Then the order of QR in each of these subsamples will be $\tau T/d$, and $\tau T/d$ is then the effective order. Next, one requires a sample of normal length, say 50, to be below the fitted quantiles for the normal laws to start to apply. This gives the simple rule $'\tau T/d \leq 50'$ for the use of the extremal inference. The rule may or may not be conservative. For example, when regressors are continuous, our computational experiments indicate that the normal inference catches up with the extremal inference when $\tau T/d \approx 15 - 20$. On the other hand, imagine having an indicator variable that picks out a 1/50-th fraction of the entire sample, as in the birthweight application, then the number of observations below the fitted quantile for this subsample will be $\tau T/50$, motivating a far more conservative rule $'\tau T/50 \leq 50'$ for this case. Overall, it seems prudent to use both extremal and normal inference at the same time in most cases, with the idea that the discrepancies between the two can indicate extreme situations.

7. Empirical Illustrations

7.1. **Extremal Risk of a Stock.**

We consider the problem of finding factors that affect value-at-risk of the Occidental Petroleum daily stock return, a problem interesting for economic analysis and real-world activities of financial firms.[9] Our dataset consists of 1000 daily observations on the one-day return Y on the Occidental Petroleum stock, covering a period in 1996-1998, and regressors, consisting of the lagged return on the spot price of oil, denoted X_1, the lagged one-day return of Dow Jones Industrials index, X_2, and the lagged own return, X_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 had been given in Chernozhukov and Umantsev (2001), and here we focus on the extremal analysis in order to illustrate new inferential tools. Figure 4 plots QR estimates $\widehat{\beta}(\tau) = (\widehat{\beta}_j(\tau), j = 0, ..., 3)$, and the shaded area represents pointwise 90% confidence intervals (CI). Normal CI are used for central quantiles, $.15 < \tau < .85$, and extremal CI are used for extremal quantiles, $\tau \leq .15$ and $\tau \geq .85$.

The coefficient on the spot price of oil is positive and increasing in the right tail of the distribution. The coefficient is large but only marginally significant at the conventional 90% level in the far right tail. However, extremal CI indicate that the distribution of the

---

[9] See Chernozhukov and Umantsev (2001), Christoffersen, Hahn, and Inoue (1999), Engle and Manganelli (1999), Diebold, Schuermann, and Stroughair (2000).

QR statistic is asymmetric, with overwhelming part of this distribution positive. Therefore, the economic effect of this determinant is potentially quite strong in the upper tail. The coefficient on the DJ index is significantly positive for the entire middle part of the distribution. The coefficient on lagged own return Y_{t-1} is significantly negative for nearly the entire distribution except for the tails, which suggests a reversion effect in the center of the distribution and no reversion effects in the tails. The lag is thus more important for the determination of intermediate risks.

We would also like to compare the extremal inference with the normal inference. This empirical example is closely matched by the Monte Carlo design with heavy-tailed $t(3)$ disturbances and regressors having the same distribution by construction. Hence we might expect that normal CI would understate the estimation uncertainty of QR and be considerably more narrow than extremal CI in the tails. As shown in Figures 5 and 6, normal CI are indeed much more narrow than extremal CI at $\tau < .1$.

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

We focus our analysis on black mothers only, motivated by Figure 7 that shows a troubling heavy tail of low birthweigts to black mothers. In our analysis we adopt the regression specification similar to that in Koenker and Hallock (2001), which has been tested for misspecification and rigorously validated for the normal quantiles by He et. al. . The specification we estimate is linear in parameters, with the response variable being the birthweight recorded in grams, and the following regressors: 'Boy' and 'Married' are indicators of infant gender and of whether the mother was married or not. 'No Prenatal', 'Prenatal Second', and 'Prenatal Third' are indicator variables that divide the sample into 4 categories: those with no prenatal visit (less than 1% of the sample), those whose first prenatal visit was in the second trimester, and those whose first prenatal visit was in the third trimester. The control category is mothers with a first visit in the first trimester, which constitute 83% of the sample. 'Smoker' and 'Cigarettes/Day' are an indicator of whether the mother smoked during pregnancy and the mother's reported average number of cigarettes smoked per day. 'Weight Gain' and 'Weight Gain^2' are the mother's reported weight gain during pregnancy (in pounds) minus the mean mother's weight gain and the mother's weight gain squared minus the square of the mean of the mother's weight gain. 'Education' is a categorical variable taking a value of 0, if the mother had less than a high-school education, 1, if completed high school education, 2, if obtained some college education, and 3, if graduated from college. 'Age' and 'Age^2' are the mother's age minus the mean of the mother's age and the mother's age squared minus the square of the mean of the mother's age.

Thus the control group consists of mothers of average age that had their first prenatal visit during the first trimester, that had attained education level of less than high school, that did not smoke, and that had average weight gain during pregnancy. The intercept in the estimated quantile regression model will measure quantiles in this group, and will therefore be called the centercept.

Figure 8 reports estimation results for extremal quantiles, and Figure 9 – for the normal quantiles. These figures show point estimates, extremal CI, and normal CI, at the confidence level of 90%. 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. Estimation results for some covariates, e.g. squared age, are not plotted to save space.

While density of birthweights, shown in Figure 7, has a finite lower support point, it has little probability mass near that boundary. This puts us in the situation similar to the Monte-Carlo design with Weibull(4) disturbances, for which differences between central and extremal inference were not large. This is what we also observe in this empirical example. For the most part, normal CI tend to be only $0-10\%$ more narrow than extremal CI, with the exception of the coefficient on 'No Prenatal', for which normal CI are twice as narrow as extremal CI. Since only 1.9% of the mothers had no prenatal care, the sample size used to estimate this coefficient is only 635. The effective order of τ-quantile for this subsample is $\tau \times 635$, which suggests that there should be little or no discrepancy between extremal CI and central CI for the coefficient on 'No prenatal' once $\tau \geq 3-5\%$. As Figure 9 shows, there is indeed little difference between extremal CI and normal CI once $\tau \geq 3\%$.

The analysis of extremal birthweights, shown in Figure 8, reveals several surprises that we do not see in the results on normal birthweights, shown in Figure 9. Most surprisingly, smoking appears to have no negative impact on extremal quantiles, in contrast to strongly negative impact at the normal quantiles. The absence of the effect could be due to possible selection, where only mothers confident of good outcomes smoke, or it could be indeed due smoking having no or little causal effect on very extreme outcomes. This finding motivates further analysis, possibly using data sets that enable instrumentation strategies.

The prenatal medical care has a strong impact on extremal quantiles and relatively little impact on normal quantiles, especially in the middle of the distribution. The effect of ‘No-prenatal’ is non-positive and may be strongly negative in the tails, anywhere between values of 0 and -800 contained in the extremal CI. The impact of ‘Prenatal Second’ and ‘Prenatal Third’ is positive and very strong in the tails in contrast to the relatively little impact found in the middle of the distribution. This could be due to mothers confident of good outcomes choosing to have a late first prenatal visit. Alternatively, it could be that substantial examinations, that accompany the first visit, during the second and especially third trimesters help intervene to reduce the probability of extreme outcomes. Finally, the mother’s weight gain seems quite important for improving extreme birthweight quantiles, much more so than for normal quantiles.

8. Conclusion

The empirical section reported several findings, which are interesting in their own right. The new tools give us the ability to maintain that the findings hold with a given statistical confidence within the postulated model.

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. The proof will employ the Convexity Lemma, cf. Geyer (1996) and Knight (1999).[10]

STEP 1. Consider the point process $\widehat{\mathbf{N}}$ defined by $\widehat{\mathbf{N}}(A) := \sum_{t=1}^{T} 1\{(A_T U_t, X_t) \in A\}$ for Borel subsets A of $E := [0, +\infty) \times \mathbb{R}^d$. The point process $\widehat{\mathbf{N}}$ converges weakly 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. This convergence follows by noting that for any set F defined as intersection a bounded rectangle with E, one has that (a) $\lim_{T\to\infty} E\widehat{\mathbf{N}}(F) = m(F) := \int_F (x'\mathbf{c})^{1/\xi} u^{1/\xi} du dF_X(x)$, where one uses regular variation of F_u and that (b) $\lim_{T\to\infty} P[\widehat{\mathbf{N}}(F) = 0] = e^{-m(F)}$, where the latter follows by Meyer's (1973) theorem after observing that we have $T\sum_{j=2}^{\lfloor T/k \rfloor} P((A_T U_1, X_1) \in F, (A_T U_j, X_j) \in F) \leq O(T\lfloor T/k \rfloor P((A_T U_1, X_1) \in F)^2) = O(1/k)$ by C5 and have geometric strong mixing by C5. Consequently, (a) and (b) imply by Kallenberg's theorem that the weak limit is a Poisson point process $\mathbf{N}$ with intensity measure m. The points of $\mathbf{N}$ can be represented as in the statement of Theorem 1: $(J_i = (\mathcal{X}_i'\mathbf{c}) \cdot \Gamma_i^{-\xi}, \mathcal{X}_i)$, for $i = 1, 2, ...$

STEP 2. Observe that $\widetilde{Z}_T(k) := A_T(\widehat{\beta}(\tau) - \beta_r) = \arg\min_{z\in\mathbb{R}^d} \sum_{t=1}^{T} \rho_\tau\big(A_T U_t - X_t'z\big)$ [write $z := A_T(\beta - \beta_r)$.] Rearranging terms gives $\sum_{t=1}^{T} \rho_\tau\big(A_T U_t - X_t'z\big) \equiv -\tau T\bar{X}'z - \sum_{t=1}^{T} 1(A_T U_t \leq X_t z)\big(A_T U_t - X_t'z\big) + \sum_{t=1}^{T} \tau A_T U_t$. Subtract $\sum_{t=1}^{T} \tau A_T U_t$ that does not depend on z and does not affect optimization, and define $Q_T(z, k) := \sum_{t=1}^{T} \ell(A_T U_t, X_t'z) = \int \ell(u, x) d\mathbf{N}(u, x)$, where $\ell(u, v) := 1(u \leq v)(v - u)$. Since ℓ is continuous and vanishes outside a compact subset of E, $Q_T(z, k)$ is a continuous mapping of $\widehat{\mathbf{N}}$, an element of $M_p(E)$, to the real line. $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ therefore implies by the Continuous Mapping Theorem that the finite-dimensional limit of $Q_T(z, k)$ is given by $Q_\infty(z, k) := -k\mu_X'z + \int_E \ell(j, x'z) d\mathbf{N}(j, x) := -k\mu_X'z + \sum_{i=1}^{\infty} \ell(J_i, \mathcal{X}_i'z)$. Then by Convexity lemma we have that $\widetilde{Z}_T(k) \in \arg\min_{z\in\mathbb{R}^d} \widetilde{Q}_T(z, k)$ converges in distribution to $\widetilde{Z}_\infty(k) = \arg\min_{z\in\mathbb{R}^d} \widetilde{Q}_\infty(z, k)$. Uniqueness and absolute continuity of $\widetilde{Z}_\infty(k)$ are shown in Lemma 1, Appendix H.

STEP 3. By C1 $A_T\left(\beta(\tau) - \beta_r\right) \to k^{-\xi}\mathbf{c}$. Thus $A_T(\widehat{\beta}(\tau) - \beta(\tau)) \to_d \widehat{Z}_\infty(k) := \widetilde{Z}_\infty(k) + k^{-\xi}\mathbf{c}$. Then $\widehat{Z}_\infty(k) = \widetilde{Z}_\infty(k) + k^{-\xi}\mathbf{c} = \arg\min_{z\in\mathbb{R}^d}[-k\mu_X'(z - k^{-\xi}\mathbf{c}) + \sum_{i=1}^{\infty} \ell(J_i, \mathcal{X}_i'(z - k^{-\xi}\mathbf{c}))]$. Claim 2 is thus proven.

STEP 4. Similarly to step 2 it follows that $\big(\widetilde{Z}_T(mk), \widetilde{Z}_T(k)\big) \in \text{ argmin }_{(z_1', z_2')' \in \mathbb{R}^{2d}} \widetilde{Q}_T(z_1, mk) + \widetilde{Q}_T(z_2, k)$ weakly converges to $\big(\widetilde{Z}_\infty(mk), \widetilde{Z}_\infty(k)\big) = \text{ argmin }_{(z_1', z_2')' \in \mathbb{R}^{2d}} \widetilde{Q}_\infty(z_1, mk) + \widetilde{Q}_\infty(z_2, k)$.

STEP 5. By Lemma 1 in Appendix H, $\mu_X'(\widetilde{Z}_\infty(mk) - \widetilde{Z}_\infty(k)) = \widetilde{Z}_\infty(mk)_1 - \widetilde{Z}_\infty(k)_1 \neq 0$ a.s. provided $mk - k > d$. It follows by the Extended Continuous Mapping Theorem that $Z_T(k) = \sqrt{k}\widehat{Z}_T(k)/(\bar{X}'(\widetilde{Z}_T(mk) - \widetilde{Z}_T(k))) \to_d Z_\infty(k) = \widehat{Z}_\infty(k)/(\mu_X'(\widetilde{Z}_\infty(mk) - \widetilde{Z}_\infty(k)))$. Claim 1 follows by using relations $\widehat{Z}_\infty(k) = \widetilde{Z}_\infty(k) + k^{-\xi}\mathbf{c}$ and $\widehat{Z}_\infty(mk) = \widetilde{Z}_\infty(mk) + (mk)^{-\xi}\mathbf{c}$ and using $\mu_X'\mathbf{c} = 1$ by C1.

---

[10] The lemma states: Suppose (i) a sequence of convex lower-semicontinous function $\widetilde{Q}_T : \mathbb{R}^d \to \bar{\mathbb{R}}$ converges in finite-dimensional sense to $Q_\infty : \mathbb{R}^d \to \bar{\mathbb{R}}$ over a dense subset of $\mathbb{R}^d$, (ii) Q_∞ is finite over a non-empty open set $\mathcal{Z}_0$, and (iii) Q_∞ is uniquely minimized at a random vector Z_∞. Then any argmin of $\widetilde{Q}_T$, denoted $\widehat{Z}_T$, converges in distribution to Z_∞.

Appendix B. Proof of Theorem 2 and 3

The results follows by Theorem 1 and the definition of convergence in distribution. □

Appendix C. Proof of Theorem 4 and 5

We will prove Theorem 4. Proof of Theorem 5 follows similarly.

The main steps of the proof, steps 1-2, are specific to our problem. In the proof we will replace $\bar{X}$ by its limit μ_X; it is easy to show that such a replacement is allowed. Let $G_T(x) := Pr\{V_T \leq x\}$ and $G(x) := \limsup_{T\to\infty} G_T(x)$.

Step 1. Define

$$\begin{aligned}\widehat{G}_{b,T}(x) &:= B_T^{-1}\sum_{i=1}^{B_T} 1\{\widehat{V}_{i,b,T}(\tau) \leq x\}\\ &= B_T^{-1}\sum_{i=1}^{B_T} 1\Big\{\mathcal{A}_{i,b,T}\psi'\left(\widehat{\beta}_{i,b,T}(\tau_b)-\beta(\tau_b)\right)+\mathcal{A}_{i,b,T}\psi'\left(\beta(\tau_b)-\widehat{\beta}(\tau_b)\right)\leq x\Big\}.\end{aligned}$$

Define also

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

where $A_b = 1/F^{-1}(1/b)$ is the scaling rate defined in Theorem 1. Then

$$\begin{aligned}&1[V_{i,b,T}\leq x-\mathcal{A}_{i,b,T}w_T/A_b]\leq 1[\widehat{V}_{i,b,T}\leq x]\leq 1[V_{i,b,T}\leq x+\mathcal{A}_{i,b,T}w_T/A_b]\\ &\text{for all } i=1,...,B_T\ ,\end{aligned}$$

where $w_T = |A_b\psi'(\beta(\tau_b)-\widehat{\beta}(\tau_b))|$.

The principal claim is that under conditions of Theorem 3 $w_T = o_p(1)$. To verify this claim note that for $k=\tau_T T$

$$\begin{aligned}A_b\times\left(\beta(\tau_b)-\widehat{\beta}(\tau_b)\right) &= O_p\left(\frac{1}{\left[F_u^{-1}(2k/b)-F_u^{-1}(k/b)\right]}\right)\times O_p\left(\frac{\left[F_u^{-1}(2k/b)-F_u^{-1}(k/b)\right]}{\sqrt{\tau_b\cdot T}}\right)\\ &= O_p\left(\frac{1}{\sqrt{\tau_b\cdot T}}\right)=O_p\left(\sqrt{\frac{b}{kT}}\right)=o_p(1)\text{ as } b/T\to 0 \text{ and } T\to\infty.\end{aligned}$$

This follows by applying the definition $A_b := -1/F_u^{-1}(1/b)$ and the basic fact from regular variation of F_u at $-\infty$ with exponent $-1/\xi$, which implies the regular variation of $F_u^{-1}(\cdot)$ at 0 with exponent $-\xi$, which implies that for any $l \searrow 0$,

$$F_u^{-1}(l)(2^{-\xi}-1)\sim\left[F_u^{-1}(2l)-F_u^{-1}(l)\right].$$

We also note that since $\tau_b = k/b$ and since $\tau_b \times T = (k/b) \times T \to \infty$ by $T/b \to \infty$ by assumption, $\widehat{\beta}(\tau_b)$ is the intermediate rank regression quantile computed using the *full sample* of size T, so that by Theorem XX its stochastic order is given as

$$\left(\beta(\tau_b) - \widehat{\beta}(\tau_b)\right) = O_p\left(\frac{\left[F_u^{-1}(k/b) - F_u^{-1}(k/b)\right]}{\sqrt{\tau_b \cdot T}}\right). \tag{C.20}$$

This verifies (**??**) and therefore claim A for the extreme rank case.

Given the claim that $w_T = o_p(1)$, for some sequence of constants $\xi_T \searrow 0$ the following event occurs wp $\to 1$:

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

Event M_T implies

$$\dot{G}_{b,T}(x;\xi_T) \leq \widehat{G}_{b,T}(x) \leq \dot{G}_{b,T}(x;-\xi_T). \tag{C.21}$$

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

$$E[\dot{G}_{b,T}(x;\xi_T)] = P[V_b - \mathcal{A}_b\xi_T/A_b \leq x]. \tag{C.22}$$

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

$$\lim_{T\to\infty} E[\dot{G}_{b,T}(x;\xi_T)] = \lim_{b\to\infty} P[V_b - \mathcal{A}_b\xi_T/a_b \leq x] = P[Z_\infty \leq x] = G(x). \tag{C.23}$$

The statement (C.23) follows because $V_b - \frac{\mathcal{A}_b\xi_T}{A_b} = V_b + o_p(1) \to_d Z_\infty$, since by Theorem 1 $V_b \to_d Z_\infty$ and $\frac{\mathcal{A}_b\xi_T}{a_b} = O_p(1)\cdot\xi_T = O_p(1)\cdot o(1) = o_p(1)$.

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

Step 3. Finally, since event M_T occurs wp $\to 1$ and so does (C.21), by Step 2 it follows that $\widehat{G}_{b,T}(x) \to_p G(x)$. Finally, convergence of quantiles is implied by the convergence of distribution functions at continuity points, i.e. $\widehat{c}_\alpha = \widehat{G}_{b,T}^{-1}(\alpha) \to_p c_\alpha = G^{-1}(\alpha)$, provided $G^{-1}(\alpha)$ is a continuity point of $G(x)$. □

Appendix D. Proof of Theorems 6 and 7

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

The proof proceeds 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 the one used in the parametric bootstrap proofs, but the complication here is the non-standard distribution of the limit processes. The demonstration of continuity poses some difficulties dealt with by invoking epi-convergence arguments and exploiting the properties of Poisson processes.

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

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

$$\sup_{x\in\mathbf{X}} |F_X(x) - \widehat{F}_X(x)| \to_p 0, \ \widehat{\xi} \to_p \xi, \ \widehat{\mathbf{c}} \to_p \mathbf{c} \text{ as } T \to \infty. \tag{D.24}$$

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: $(\Gamma_j^{-\xi} \cdot \mathcal{X}_j' \mathbf{c}, \quad \mathcal{X}_j)$, $j = 1, 2, 3, ...$ $\mathbf{N}$ is a random element of complete and separable metric space of point measures $(M_p(E), \rho_v)$ with metric ρ_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 $\mathbf{N}$. The law of $\mathbf{N}$ in $(M_p(E), \rho_p)$ will be denoted as $\mathcal{L}(\mathbf{N}|\xi, \mathbf{c}, F_X)$. The law depends only on the parameters $\xi, \mathbf{c}, F_X$ of the intensity measure $m_{\mathbf{N}}$.

3. The random objective function (ROF) for $\mu_X = \int x dF_X(x)$ takes the form

$$\begin{aligned} Q_\infty(z;k) &= -k\mu_X'(z - k^{-\xi}\mathbf{e}_1) + \int_E (u^{-\xi} \cdot x'\mathbf{c} + \mathcal{X}_i'(z - k^{-\xi}\mathbf{c}))_+ d\mathbf{N}(u,x) \\ &= -k\mu_X'(z - k^{-\xi}\mathbf{c}) + \sum_{i=1}^{\infty} (\Gamma_i^{-\xi} \cdot \mathcal{X}_i'\mathbf{c} + \mathcal{X}_i'(z - k^{-\xi}\mathbf{c}))_+ \end{aligned} \tag{D.25}$$

and is a random element of the metric space of proper lower-semi-continuous functions $(LC(\mathbb{R}^d), \rho_e)$, where the metric ρ_e induced by the topology of *epi-convergence*. Moreover, this function is convex in z, which is very important to what follows. The underlying probability space $(\Omega', \mathcal{F}', P')$ is the one induced by Monte-Carlo draws of points of $\mathbf{N}$. The law of $Q_\infty(z;k)$ in $(LC(\mathbb{R}^d), \rho_e)$ will be denoted as $\mathcal{L}(Q_\infty(z;k)|\xi, \mathbf{c}, F_X)$. The law depends only on the parameters $\xi, \mathbf{c}, F_X$.

4. The extremum statistic $Z_\infty(z;k) = \arg\min_{z\in\mathbb{R}^d} Q_\infty(z;k)$ is a random element in the metric space $\mathbb{R}^d$, equipped with the usual Euclidian metric. The underlying probability space $(\Omega', \mathcal{F}', P')$ is the one induced by Monte-Carlo draws of points of $\mathbf{N}$. The law of $Z_\infty(z;k)$ in $\mathbb{R}^d$ will be denoted as $\mathcal{L}(Z_\infty(z;k)|\xi, \mathbf{c}, F_X)$. The law depends only on the parameters $\xi, \mathbf{c}, F_X$.

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

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

$$\lim_{m\to\infty} \rho_w(\mathcal{L}(\mathbf{N}^m|\xi^m, \mathbf{c}^m, F_X^m), \mathcal{L}(\mathbf{N}|\xi, \mathbf{c}, F_X)) = 0. \tag{D.26}$$

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

$$\lim_{m\to\infty} \chi(f; \mathbf{N}^m) = \chi(f; \mathbf{N}), \quad (\forall f \in C_K^+(E)), \tag{D.27}$$

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:

$$\chi(f;\mathbf{N}) \quad := E\left[e^{\int_E f(u,d)\ d\mathbf{N}}\right] = e^{\left(-\int_E \left[1-e^{-f(u,x)}\right] dm_{\mathbf{N}}(u,x)\right)}. \tag{D.28}$$

B. A sequence of ROFs $\{Q_\infty^m(\cdot;k),,m = 1,2,3,...\}$ defined by the sequence of parameters $\{\xi^m, \mathbf{c}^m, F_X^m, m = 1,2,3,...\}$ converges weakly to the ROF $Q_\infty(\cdot;k)$ defined by parameters $\xi, \mathbf{c}, F_X$ 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 denoted ρ_w(or any other metric that metrizes weak convergence):

$$\lim_{m\to\infty} \rho_w(\mathcal{L}(Q_\infty^m(\cdot;k)|\xi^m, \mathbf{c}^m, F_X^m), \mathcal{L}(Q_\infty(\cdot;k)|\xi, \mathbf{c}, F_X)) = 0. \tag{D.29}$$

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

$$(Q_\infty^m(z_j;k), j = 1,...,J) \to_d (Q_\infty(z_j;k), j = 1,...J) \quad \text{ in } \mathbb{R}^J \tag{D.30}$$

for any finite collection of points $(z_j, j = 1,...,J)$. This second condition is much easier to check than the previous one.

C. In turn, the weak convergence of objective functions $\{Q_\infty^m(z;k), m = 1,2,3,...\}$ to $\{Q_\infty(z;k)$ in $(LC(\mathbb{R}^d), \rho_e)$ implies that as $m \to \infty$ the weak convergence of the corresponding argmins: $Z_\infty^m(z;k) \to_d Z_\infty(z;k)$ in $\mathbb{R}^d$, that is,

$$\lim_{m\to\infty} \rho_w(\mathcal{L}(Z_\infty^m(z;k)|\xi^m, \mathbf{c}^m, F_X^m), \mathcal{L}(Z_\infty(z;k)|\xi, \mathbf{c}, F_X)) = 0. \tag{D.31}$$

The proof is now done in two steps:

I. We would like to show that the law $\mathcal{L}(Z_\infty(z;k)|\xi', \mathbf{c}', F_X')$ is continuous at $(\xi', \mathbf{c}', F_X') = (\xi, \mathbf{c}, F_X)$, that is, for any sequence $(\xi^m, \mathbf{c}^m, F_X^m, m = 1,2,...)$ such that

$$|\xi^m - \xi| \to 0, |\mathbf{c}^m - \mathbf{c}| \to 0, \sup_{x\in\mathbf{X}} |F_X^m(x) - F_X(x)| \to 0 \tag{D.32}$$

we have

$$\rho_w(\mathcal{L}(Z_\infty(z;k)|\xi^m, \mathbf{c}^m, F_X^m), \mathcal{L}(Z_\infty(z;k)|\xi, \mathbf{c}, F_X)) \to 0. \tag{D.33}$$

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

$$\rho_w(\mathcal{L}(Z_\infty(z;k)|\widehat{\xi}, \widehat{\mathbf{c}}, \widehat{F}_X), \mathcal{L}(Z_\infty(z;k)|\xi, \mathbf{c}, F_X)) \to_p 0. \tag{D.34}$$

That is, the law $\mathcal{L}(Z_\infty(z;k)|\widehat{\xi}, \widehat{\mathbf{c}}, \widehat{F}_X)$ generated by the Monte Carlo procedure consistently estimates the limit law $\mathcal{L}(Z_\infty(z;k)|\xi, \mathbf{c}, F_X)$, which is what we needed to prove, since the above result implies that $\rho_L(\mathcal{L}(\psi' Z_\infty(z;k)|\widehat{\xi}, \widehat{\mathbf{c}}, \widehat{F}_X), \mathcal{L}(\psi' Z_\infty(z;k)|\xi, \mathbf{c}, F_X)) \to_p 0$, which implies 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.

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

$$(D.32) \overset{(1)}{\Longrightarrow} (D.27) \overset{(2)}{\Longrightarrow} (D.26) \quad \overset{(3)}{\Longrightarrow} (D.30) \overset{(4)}{\Longrightarrow} (D.29) \overset{(5)}{\Longrightarrow} (D.31) \Leftrightarrow (D.33) \qquad \text{(D.35)}$$

where (1) follows by direct calculations:

$$|\chi(f;\mathbf{N}^m) - \chi(f;\mathbf{N})| \leq \chi(f;\mathbf{N})|e^{\int_K g(u,x)dm_{\mathbf{N}} - \int_K g(u,x)dm_{\mathbf{N}^m}} - 1| \to 0, \qquad \text{(D.36)}$$

as $\int_K (g(u,x)\,[dm_{\mathbf{N}} - dm_{\mathbf{N}^m}]) \to 0$, which follows from the definition of the measure $m_{\mathbf{N}}$ stated earlier; (2) by the preceding discussion in Step A, (3) by the continuity of the mapping $\mathbf{N} \mapsto \int_E (u^{-\xi} \cdot x'\mathbf{c} + (z - k^{-\xi}))_+ d\mathbf{N}(u,x)$ from $(M_p(E), \rho_v)$ to $\mathbb{R}$, (4) and (5) by the preceding discussion in Step C. ☐

Appendix E. Uniqueness and Continuity

Define $k := \lim_{T\to\infty} \tau T$. Let $\{\mathcal{X}_t, t \geq 1\}$ be an i.i.d. sequence from distribution function F_X such that $E[\mathcal{X}\mathcal{X}']$ is positive definite and $P_{F_X}\{G_k \in \partial(0,1)^d\} = 0$, where $G_k := (k\mu_X - \sum_{t\leq\lfloor k\rfloor} \mathcal{X}_t)[X_{\lfloor k\rfloor+1} ... X_{\lfloor k\rfloor+d}]^{-1'}$, if the matrix $[X_{\lfloor k\rfloor+1} ... X_{\lfloor k\rfloor+d}]^{-1'}$ is invertible and $G := (\infty, ..., \infty)$, otherwise. Denote by $\mathcal{F}_X(k)$ the class of distributions F_X for which the above property holds. Denote by $\mathcal{F}_X$ the class of distributions F_X that belong to $\mathcal{F}_X(k')$ for both $k' = k$ and $k' = mk$.

Lemma 1. *If $F_X \in \mathcal{F}_X$, then $\widehat{Z}_\infty(k)$ and $Z_\infty(k)$, defined in Theorem 1, are uniquely defined random vectors that have an absolutely continuous distribution.*

Comment E.1. The stated condition is an analog of Koenker and Bassett's (1978) condition for uniqueness in finite samples. This condition trivially holds if for a given k 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: We have from Theorem 1 that $\widehat{Z}_\infty(k) = \widetilde{Z}_\infty(k) +$const. Chernozhukov (2005) shows that a sufficient condition for tightness of possibly set-valued $\widetilde{Z}_\infty(k)$ is $E[\mathcal{X}\mathcal{X}'] > 0$. Taking tightness as given, conditions for uniqueness and continuity of $\widetilde{Z}_\infty(k)$ can be established. Define $\mathcal{H}$ as the set of all d-element subsets of $\mathbb{N} = \{1, 2, 3, ...\}$. 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. Let $\mathcal{H}^* = \{h \in \mathcal{H} : |\mathcal{X}(h)| \neq 0\}$. $\mathcal{H}^*$ is non-empty a.s. by C2 and is countable. Application of the argument of Theorem 3.1. of Koenker and Bassett (1978) gives that at least one element of $\widetilde{Z}_\infty(k)$ takes the form $z_h = \mathcal{X}(h)^{-1}J(h)$ for some $h \in \mathcal{H}^*$, and must satisfy the gradient condition: $\zeta_k(z_h) := (k\mu_X - \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}_t' z_h)\mathcal{X}_t)'\mathcal{X}(h)^{-1} \in [0,1]^d$, where the argmin is unique if and only if $\zeta_k(z_h) \in \mathcal{D} = (0,1)^d$. Thus, uniqueness holds for a fixed $k > 0$ if $P(\exists h \in \mathcal{H}^* : \zeta_k(z_h) \in \partial\mathcal{D}) = 0$. To bound the latter probability, define $\mathcal{M}(k)$ as the set of all $\lfloor k\rfloor$-element subsets of $\mathbb{N}$. Then note that the latter probability is bounded by

$$P_{F_X}\{G(m,h) \in \partial\mathcal{D}, \exists h \in \mathcal{H}, \exists m \in \mathcal{M}(k) : h \cap m = \emptyset\} \leq \sum_{h\in\mathcal{H}, m\in\mathcal{M}: h\cap m=\emptyset} P_{F_X}\{G(m,h) \in \partial\mathcal{D}\},$$

where $G(m,h) := (k\mu_X - \sum_{t\in m} \mathcal{X}_t)'\mathcal{X}(h)^{-1}$, if $\mathcal{X}(h)$ is invertible, and $G(m,h) := (\infty, ..., \infty)$, if $\mathcal{X}(h)$ is not invertible. Thus $P_{F_X}\{G(m,h) \in \partial\mathcal{D}\} = 0$ for any $h \in \mathcal{H}$ and $m \in \mathcal{M}$ such that $h \cap m = \emptyset$ is the condition that suffices for uniqueness. By the i.i.d. assumption and $h \cap m = \emptyset$, the latter probability in fact equals $P_{F_X}\{G_k \in \partial\mathcal{D}\}$, where $G_k := (k\mu_X - \sum_{t \le \lfloor k \rfloor} \mathcal{X}_t)'[X_{\lfloor k \rfloor+1}...X_{\lfloor k \rfloor+d}]^{-1'}$, if the matrix $[X_{\lfloor k \rfloor+1}...X_{\lfloor k \rfloor+d}]^{-1'}$ is invertible and $G_k := (\infty, ..., \infty)$ otherwise. Therefore, $P_{F_X}\{G_k \in \partial\mathcal{D}\} = 0$ is the condition that suffices for uniqueness.

Using the above uniqueness property, existence of the density of $\widetilde{Z}_\infty(k)$ follows by the argument of Theorem 3.4 in Koenker and Bassett (1978) used to derive the finite-sample density of QR. Given $\{\mathcal{X}_t\}$, $h \in \mathcal{H}^*$, and $J(h)$, the probability that $\widetilde{Z}_\infty(k) = \mathcal{X}(h)^{-1}J(h)$ equals $P\big\{\zeta_k(\mathcal{X}(h)^{-1}J(h)) \in \mathcal{D}\big|\{\mathcal{X}_t\}, J(h)\big\}$. Conditional on $\{\widetilde{Z}_\infty(k) = \mathcal{X}(h)^{-1}J(h)\}$, $h \in \mathcal{H}^*$, and $\mathcal{X}(h)$, the density of $\widetilde{Z}_\infty(k)$ at z is $f_{J(h)|\mathcal{X}(h)}(\mathcal{X}(h)z) \cdot |\mathcal{X}(h)|$, where $f_{J(h)|\mathcal{X}(h)}(u)$ is the joint density of $J(h)$ conditional on $\mathcal{X}(h)$. Thus, the joint density of $Z_\infty(k)$ at z is $f_{Z_\infty(k)}(z) = E[\sum_{h \in \mathcal{H}^*} f_{J(h)|\mathcal{X}(h)}(\mathcal{X}(h)z) \cdot |\mathcal{X}(h)| \cdot P\{\zeta_k(\mathcal{X}(h)^{-1}J(h)) \in \mathcal{D}|\{\mathcal{X}_t\}, J(h)\}]$.

We have that $Z_\infty(k) = \widetilde{Z}_\infty(k)/(\mu_X'(\widetilde{Z}_\infty(mk) - \widetilde{Z}_\infty(k))$. By the argument identical to that in the proof Theorem 2.2 in Bassett and Koenker (1981) we a.s. have that (a) $\widetilde{Z}_\infty(mk) = \mathcal{X}(h_1)^{-1}J(h_1)$ and $\widetilde{Z}_\infty(k) = \mathcal{X}(h_1)^{-1}J(h_2)$ for some $h_1 \neq h_2 \in \mathcal{H}^*$, and (b) $\mu_X'(\widetilde{Z}_\infty(mk) - \widetilde{Z}_\infty(k)) = (\widetilde{Z}_\infty(mk)_1 - \widetilde{Z}_\infty(k))_1 = (\mathcal{X}(h_1)^{-1}J(h_1) - \mathcal{X}(h_2)^{-1}J(h_2))_1 > 0$ a.s.. Since the first column of $\mathcal{X}(h)$ is a vector of ones and $\mathcal{X}(h)$ is invertible for $h \in \mathcal{H}^*$, all the entries of the first row of $\mathcal{X}(h)^{-1}$ are non-zero. Therefore, $(\mathcal{X}(h_1)^{-1}J(h_1) - \mathcal{X}(h_2)^{-1}J(h_2))_1$ is necessarily a linear function of both $J(h_1)$ and $J(h_2)$ with non-zero coefficients. This property and $h_1 \neq h_2$ imply that, conditional on regressors, $\mathcal{X}(h_1)^{-1}J(h_1)$ and $(\mathcal{X}(h_1)^{-1}J(h_1) - \mathcal{X}(h_2)^{-1}J(h_2))_1$ are jointly absolutely continuous in $\mathbb{R}^{d+1}$. By repeating the arguments in the previous paragraph, it then follows that $\mu_X'(\widetilde{Z}_\infty(mk) - \widetilde{Z}_\infty(k))$ and $\widetilde{Z}_\infty(k)$ are jointly absolutely continuous. By a transformation argument, it then follows that $Z_\infty(k)$ is a well-defined absolutely continuous vector.

References

Abreveya, J. (2001): "The effects of Demographics and Maternal Behavior on the Distribution of borth outcomes," *Empirical Economics*, 26(1), 247–257.

Aigner, D. J., T. Amemiya, and D. J. Poirier (1976): "On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function," *Internat. Econom. Rev.*, 17(2), 377–396.

Arrow, K., T. Harris, and J. Marschak (1951): "Optimal Inventory Policy," *Econometrica*, 19, 205–272.

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.

——— (2000a): "Conditional Extremes and Near-Extremes: Estimation, Inference, and Economic Applications," Ph.D. Dissertation, Stanford (www.mit.edu/~vchern).

——— (2000b): "Extremal Quantile Regression (Linear Models and an Asymptotic Theory)," MIT, Department of Economics, Technical Report, forthcoming in the Annals of Statistics, download at `www.mit.edu/~vchern`.

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.

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.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling extremal events*, vol. 33 of *Applications of Mathematics*. Springer-Verlag, Berlin, For insurance and finance.

Feigin, P. D., and S. I. Resnick (1994): "Limit distributions for linear programming time series estimators," *Stochastic Process. Appl.*, 51(1), 135–165.

Geyer, C. J. (1996): "On the asymptotics of convex stochastic optimization," Technical Report, University of Minnesota, Department of Statistics.

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., 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.

Meyer, R. M. (1973): "A poisson-type limit theorem for mixing sequences of dependent 'rare' events," *Ann. Probability*, 1, 480–483.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *J. Econometrics*, 79(1), 147–168.

Politis, D. N., and J. P. Romano (1994): "Large sample confidence regions based on subsamples under minimal assumptions," *Ann. Statist.*, 22(4), 2031–2050.

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.

Roy, A. D. (1952): "Safety First and the Holding of Assets," *Econometrica*, 20, 431–449.

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.

Appendix F. Figures

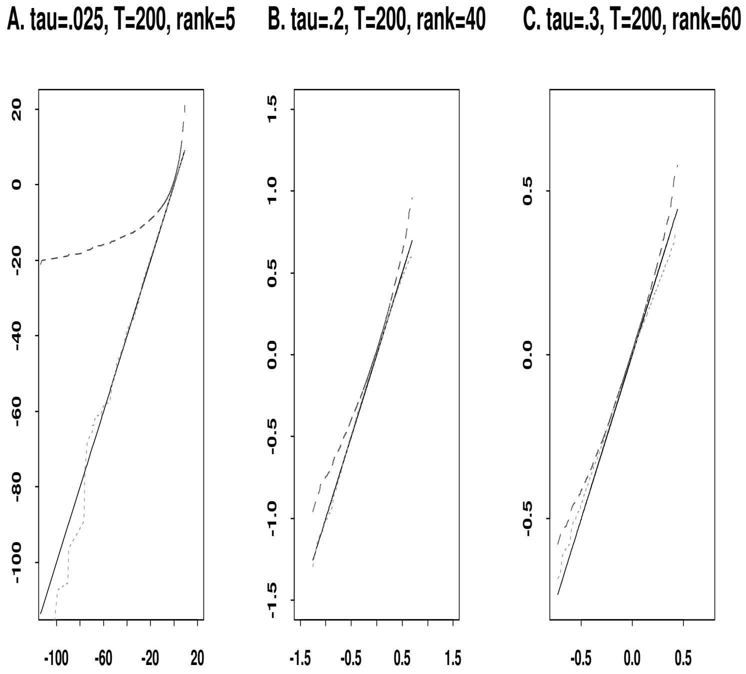

Figure 1. **T=200. Quantile vs. Quantile (QQ) plot of Extremal and Normal Approximations**. The figure is based on the simples design with $Y = X + U$, where U follows Cauchy distribution and $X = 1$. The dashed line "- - -" is the normal approximation, and the dotted line "......" is the extremal approximation. The actual quantiles of the exact sampling distribution are depicted by the solid line "——". The normal approximation becomes comparable to the extremal approximation only at $\tau T = 60$, which corresponds to $\tau = .3$ when $T = 200$. Replications $= 10,000$. QQ plots are over the 99% range.

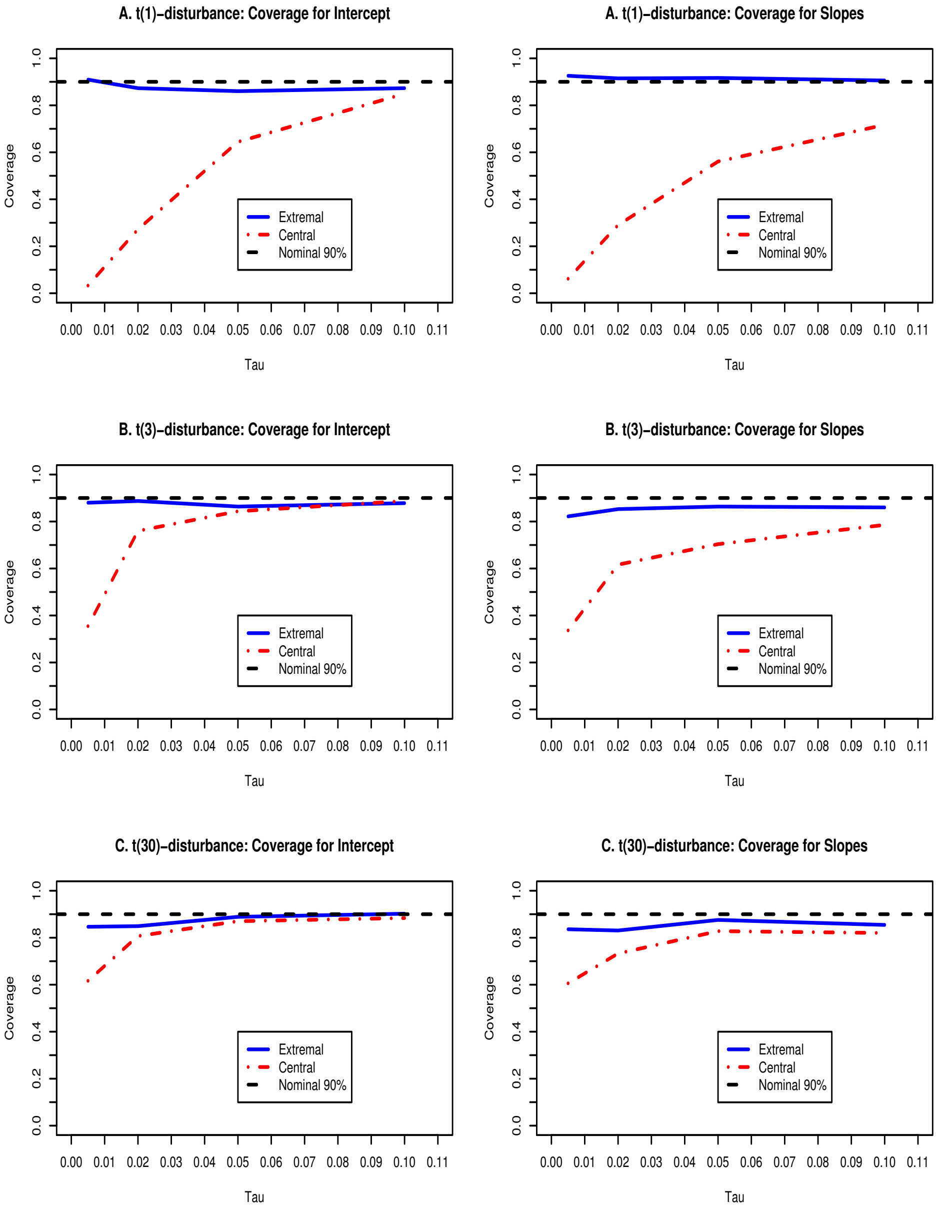

FIGURE 2. Coverage of Extremal Confidence Intervals and Normal Confidence Intervals when Disturbances are $t(\nu), \nu \in \{1, 3, 30\}$.

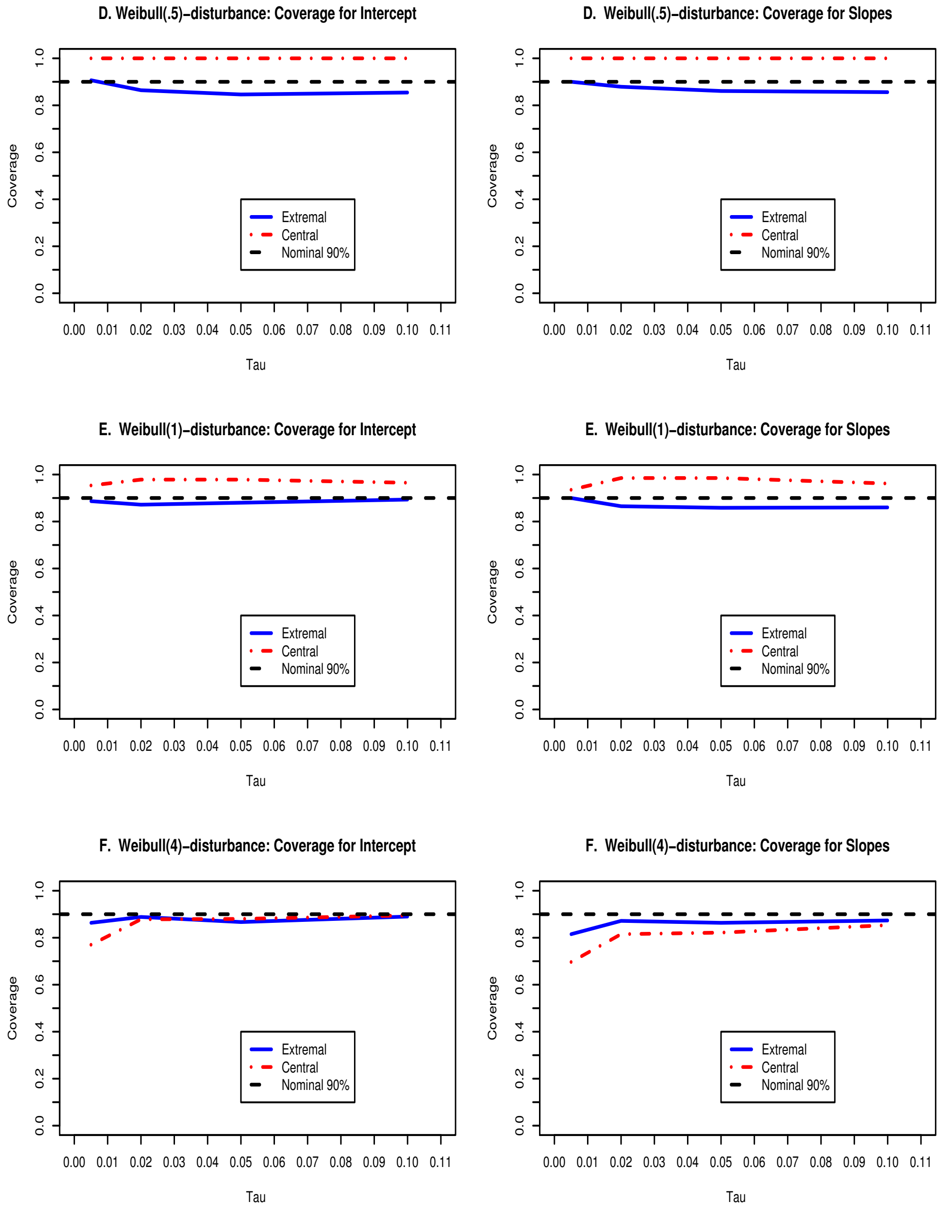

Figure 3. Coverage of Extremal Confidence Intervals and Normal Confidence Intervals when Disturbances are Weibull $(\alpha), \alpha \in \{.5, 1, 4\}$.

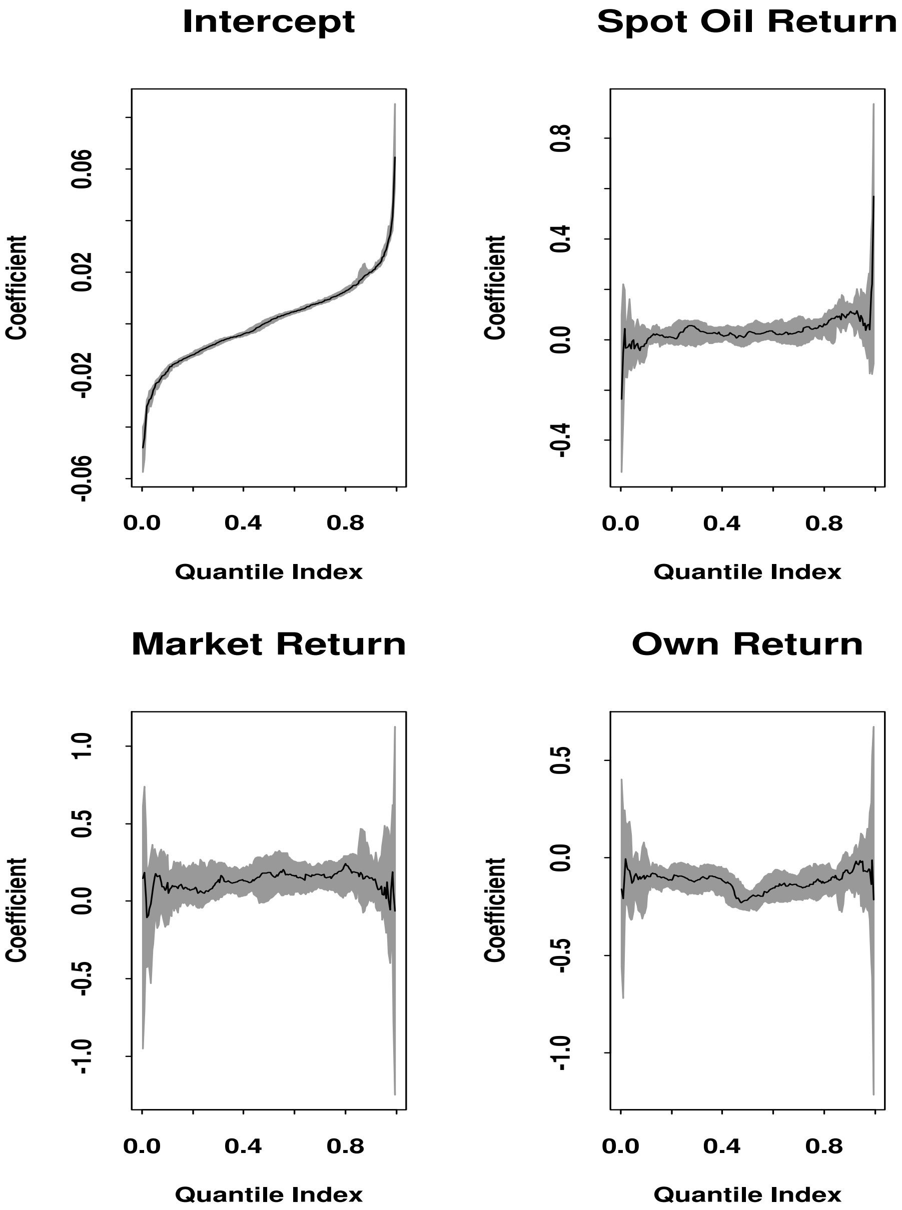

Figure 4. Regression Quantile Coefficient Estimates and 90% confidence intervals. For the central quantiles, $.15 < \tau < .85$, inference is implemented using the normal approach. For the extreme quantiles, $\tau \leq .15$ and $\tau > .85$, inference is implemented using the extremal approach.

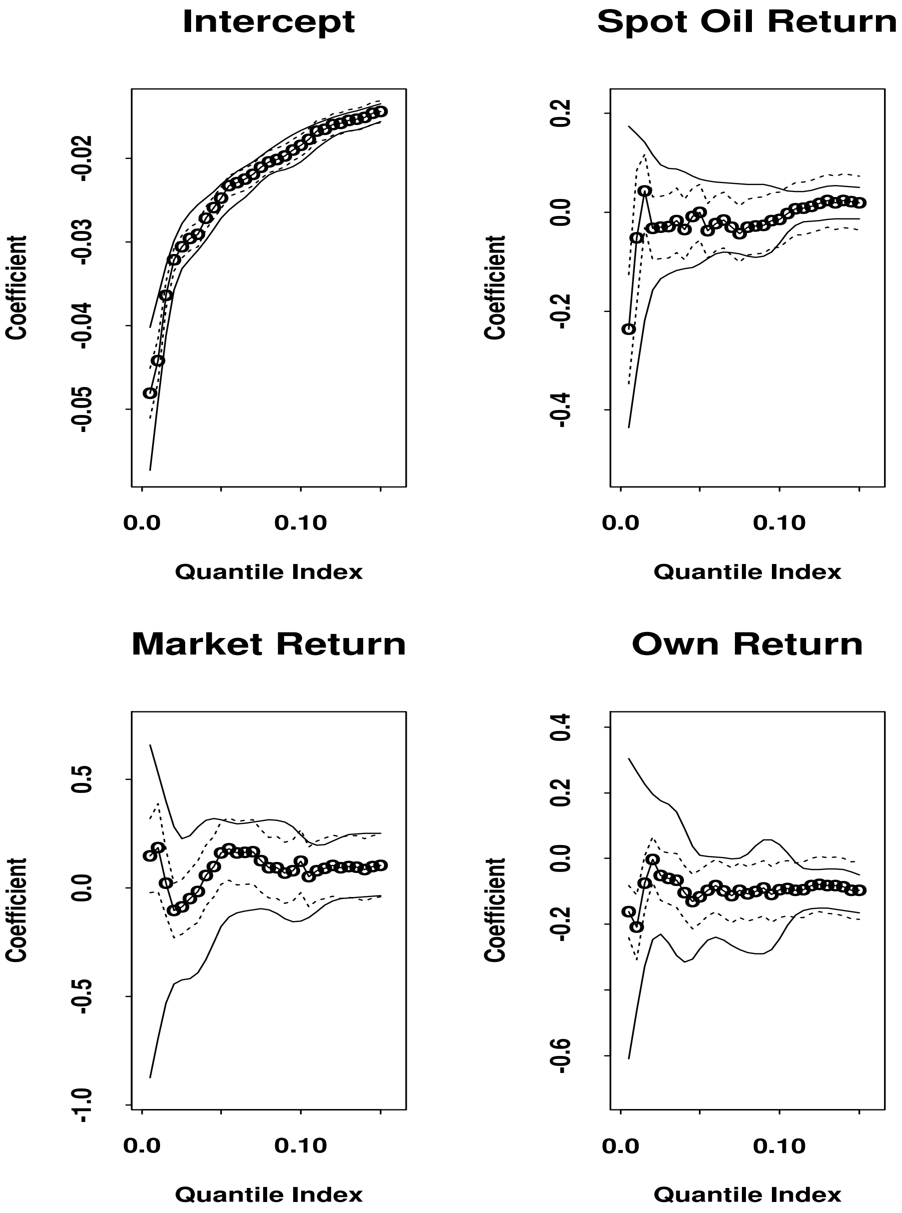

Figure 5. Regression Quantile Coefficient Estimates and 90% confidence intervals for $\tau \leq .15$. The solid line depicts the extremal confidence intervals. The dashed line depicts the normal confidence intervals.

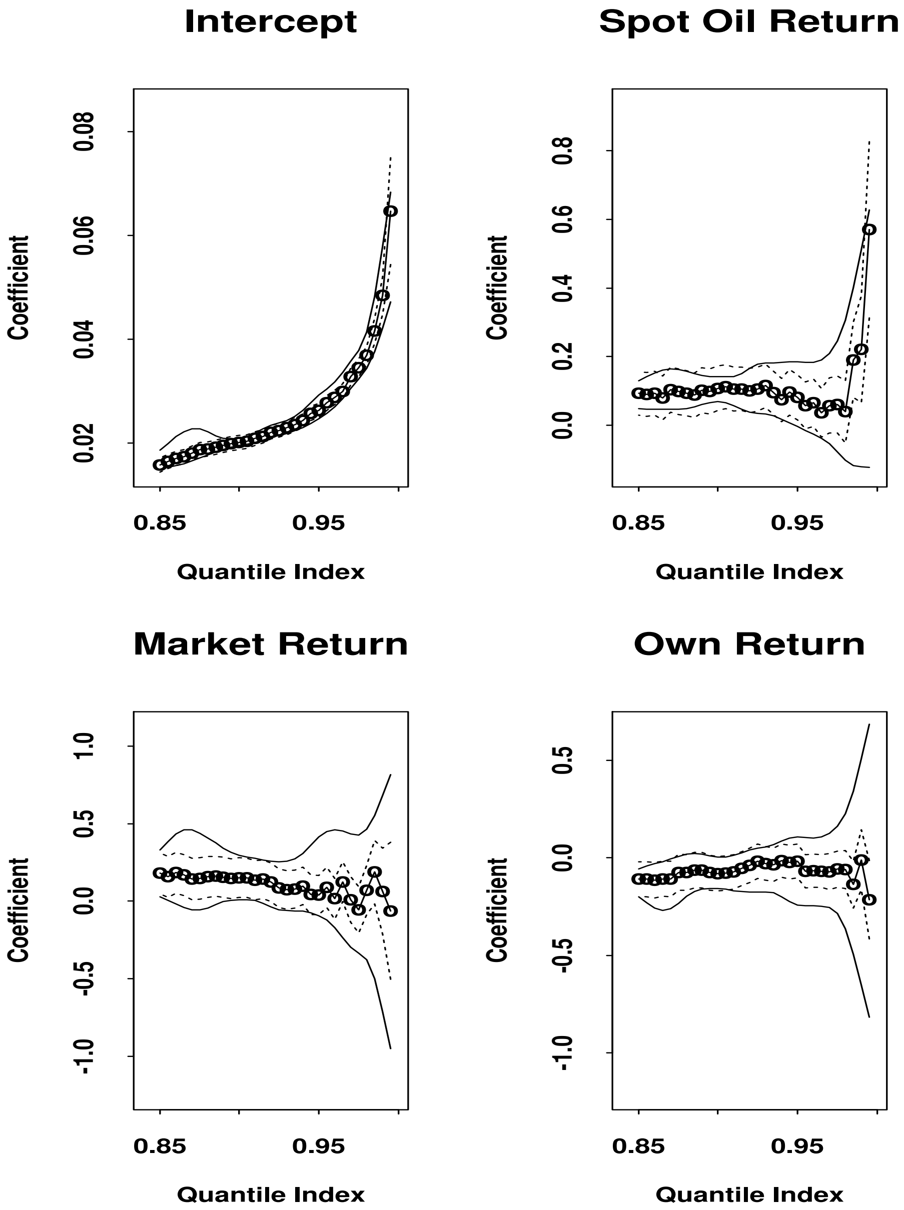

FIGURE 6. Regression Quantile Coefficient Estimates and 90% intervals for $\tau \geq .85$. The solid line depicts the extremal confidence intervals. The dashed line depicts the normal confidence intervals.

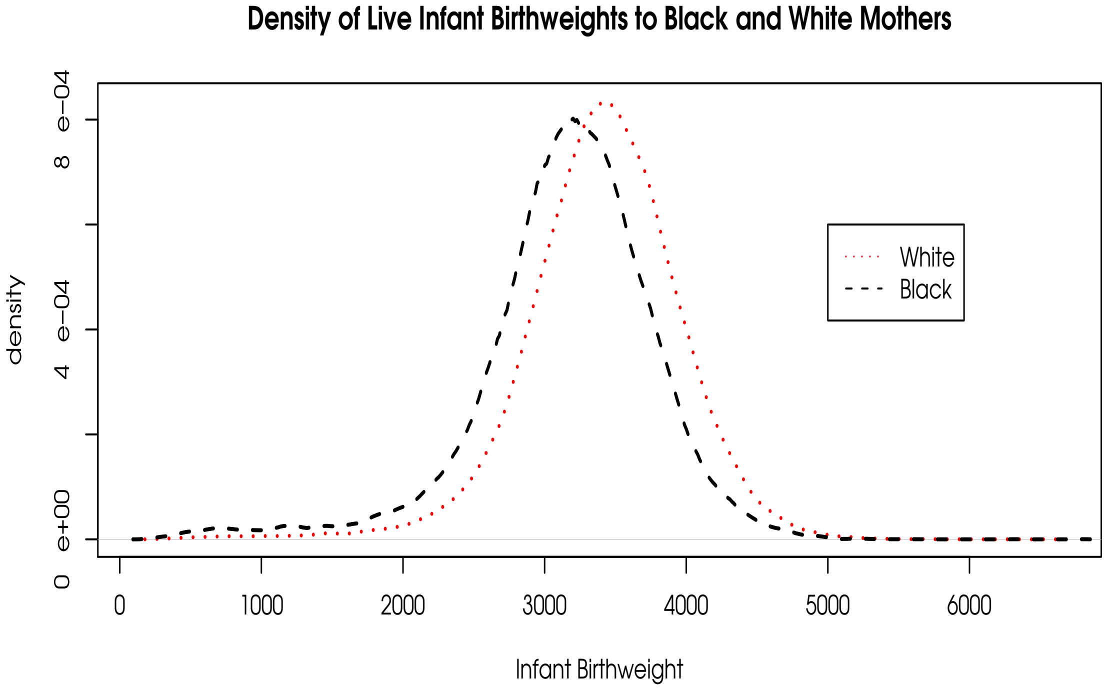

Figure 7. Birthweight Densities

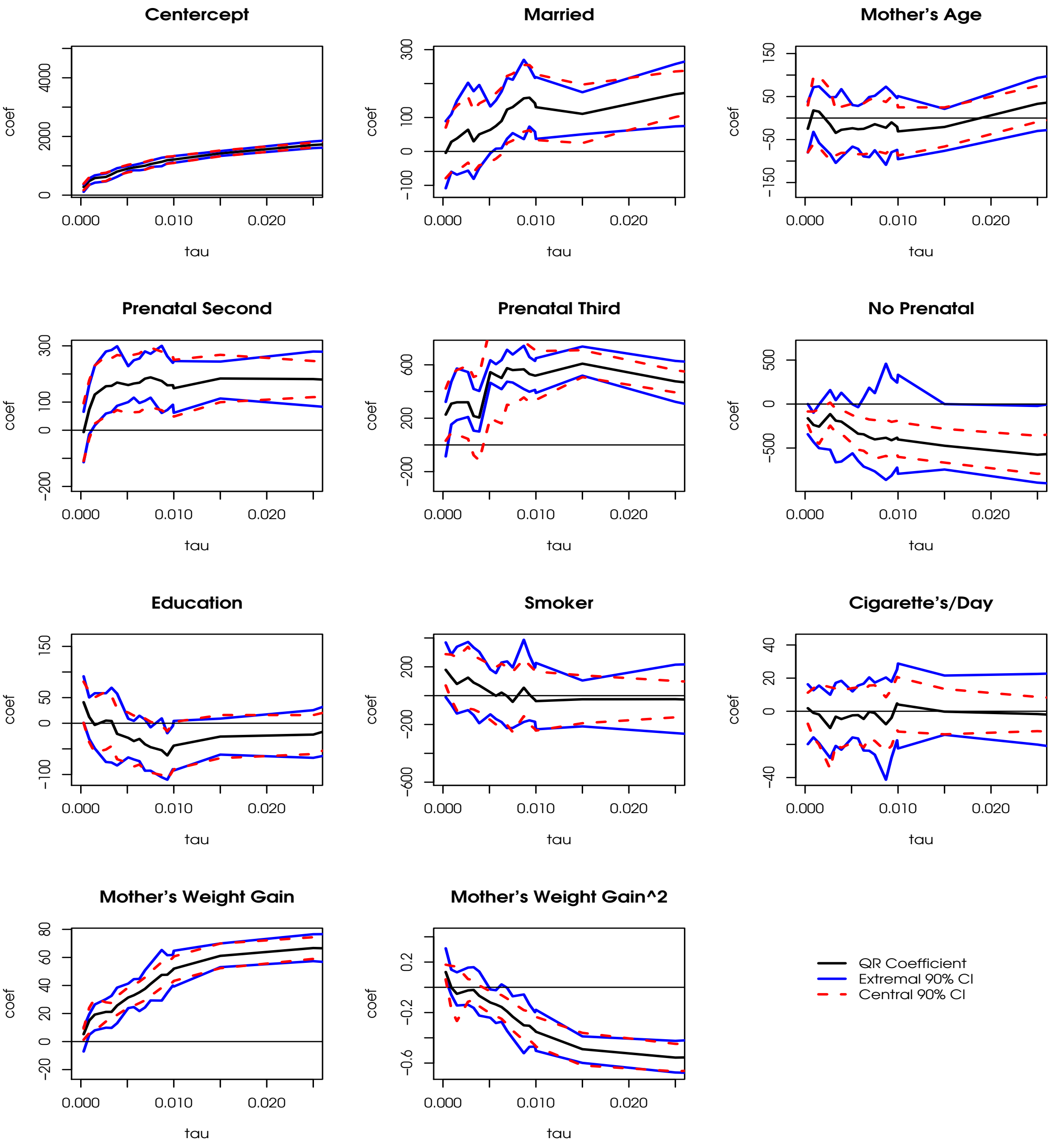

FIGURE 8. Regression Quantile Coefficient Estimates and 90% intervals for $\tau \leq .025$. The solid line depicts the extremal confidence intervals. The dashed line depicts the normal confidence intervals.

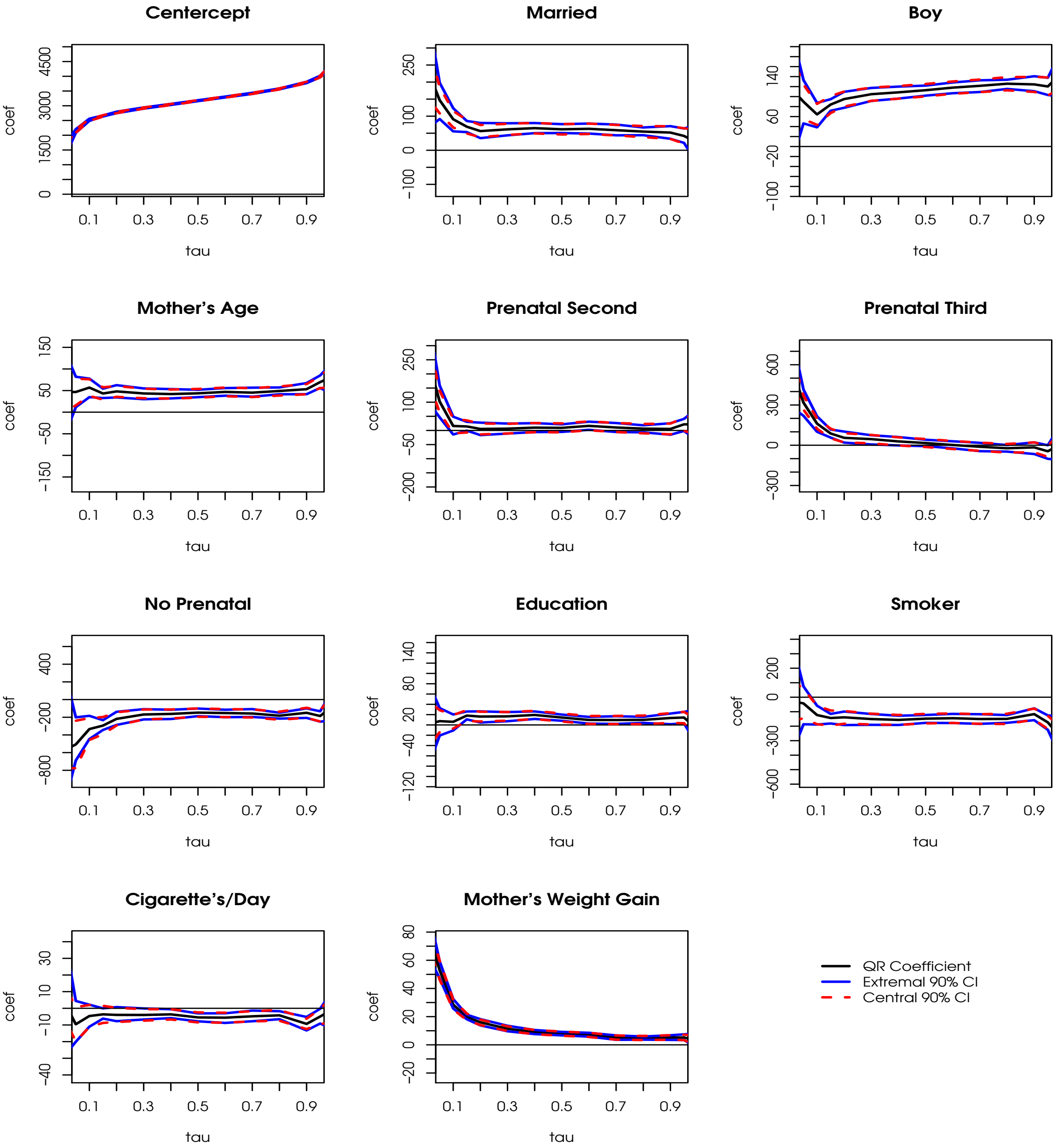

Figure 9. Regression Quantile Coefficient Estimates and 90% intervals for $\tau \in [.025, .975]$. The solid line depicts the extremal confidence intervals. The dashed line depicts the normal confidence intervals.