The Annals of Statistics
2005, Vol. 33, No. 2, 806–839
DOI 10.1214/009053604000001165

EXTREMAL QUANTILE REGRESSION[1]

BY VICTOR CHERNOZHUKOV

Massachusetts Institute of Technology

Quantile regression is an important tool for estimation of conditional quantiles of a response Y given a vector of covariates X. 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. This paper develops a theory of quantile regression in the tails. Specifically, it obtains the large sample properties of extremal (extreme order and intermediate order) quantile regression estimators for the linear quantile regression model with the tails restricted to the domain of minimum attraction and closed under tail equivalence across regressor values. This modeling setup combines restrictions of extreme value theory with leading homoscedastic and heteroscedastic linear specifications of regression analysis. In large samples, extreme order regression quantiles converge weakly to $\arg\min$ functionals of stochastic integrals of Poisson processes that depend on regressors, while intermediate regression quantiles and their functionals converge to normal vectors with variance matrices dependent on the tail parameters and the regressor design.

1. Introduction. Regression quantiles [Koenker and Bassett (1978)] estimate conditional quantiles of a response variable Y given regressors X. They extend Laplace's (1818) median regression (least absolute deviation estimator) and generalize the ordinary sample quantiles to the regression setting. Regression quantiles are used widely in empirical work and studied extensively in theoretical statistics. See, for example, Buchinsky (1994), Chamberlain (1994), Chaudhuri, Doksum and Samarov (1997), Gutenbrunner and Jurečková (1992), Hendricks and Koenker (1992), Knight (1998), Koenker and Portnoy (1987), Portnoy and Koenker (1997), Portnoy (1991a) and Powell (1986), among others.

Many potentially important applications of regression quantiles involve the study of various extremal phenomena. In econometrics, motivating examples include the analysis of factors that contribute to extremely low infant birthweights [cf. Abrevaya (2001)]; the analysis of the highest bids in auctions [cf. Donald and Paarsch (1993)]; and estimation of factors of high risk in finance [cf. Tsay (2002) and Chernozhukov and Umantsev (2001), among others]. In biostatistics and other areas, motivating examples include the analysis of survival at extreme durations [cf. Koenker and Geling (2001)]; the analysis of factors that impact

Received August 2003; revised July 2004.
[1]Supported in part by an Alfred P. Sloan Foundation Dissertation Fellowship.
AMS 2000 subject classifications. Primary 62G32, 62G30, 62P20; secondary 62E30, 62J05.
Key words and phrases. Conditional quantile estimation, regression, extreme value theory.

the approximate boundaries of biological processes [cf. Cade (2003)]; image reconstruction and other problems where conditional quantiles near maximum or minimum are of interest [cf. Korostelëv, Simar and Tsybakov (1995)].

An important peril to inference in the listed examples is that conventional large sample theory for quantile regression does not apply sufficiently far in the tails. In the nonregression case, this problem is familiar, well documented and successfully dealt with by modern extreme value theory; see, for example, Leadbetter, Lindgren and Rootzén (1983), Resnick (1987) and Embrechts, Klüppelberg and Mikosch (1997). The purpose of this paper is to develop an asymptotic theory for quantile regression in the tails based on this theory. Specifically, this paper obtains the large sample properties of extremal (extreme order and intermediate order) quantile regression for the class of linear quantile regression models with conditional tails of the response variable restricted to the domain of minimum attraction and closed under the tail equivalence across conditioning values.

The paper is organized as follows. After an introductory Section 2, Section 3 joins together the linear quantile regression model with the tail restrictions of modern extreme value theory. These restrictions are imposed in a manner that allows regressors to impact the conditional tail quantiles of response Y differently than the central quantiles. The resulting modeling setup thus covers conventional location shift regression models, as well as more general quantile regression models. Section 4 provides the asymptotic theory for the sample regression quantiles under the extreme order condition, $\tau_T T \to k > 0$, where τ_T is the quantile index and T is the sample size. By analogy with the extreme order quantiles in nonregression cases, the extreme order regression quantiles converge to extreme type variates (functionals of multivariate Poisson processes that depend on regressors). Our analysis of the case $\tau_T T \to k > 0$ builds on and complements the analysis of $\tau_T T \to 0$ given by Feigin and Resnick (1994), Smith (1994), Portnoy and Jurečková (1999) and Knight (2001) for various types of location shift models. [Chernozhukov (1998) also studied some nonparametric cases.] Section 5 derives the asymptotic distributions of regression quantiles under the intermediate order condition: $\tau_T T \to \infty$, $\tau_T \to 0$, thus providing a quantile regression analog of the results on the intermediate univariate quantiles by Dekkers and de Haan (1989). As with the intermediate quantiles in nonregression cases, the intermediate order regression quantiles, and their functionals such as Pickands type estimators of the extreme value index, analyzed in Section 6, are asymptotically normal with variance determined by both the tail parameters and the regressor design. Section 7 provides an illustration, Section 8 concludes, and Section 9 collects the proofs.

2. The setting.

Suppose Y is the response variable in $\mathbb{R}$, and $X = (1, X'_{-1})'$ is a $d \times 1$ vector of regressors (typically transformations of original regressors). (Throughout the paper, given a vector x, x_{-1} denotes x without its first component x_1.) Denote the conditional distribution of Y given $X = x$ by $F_Y(\cdot|x)$.

The present focus is on $F_Y^{-1}(\tau|x) = \inf\{y : F_Y(y|x) > \tau\}$, where τ is close to 0. Let there be a sample

$$\{Y_t, X_t, t = 1, \ldots, T\} \qquad \text{where } X_t \in \mathbf{X},$$

generated by a probability model with a conditional quantile function of the classical linear-in-parameter form

$$F_Y^{-1}(\tau|x) = x'\beta(\tau) \qquad \text{for all } \tau \in \mathcal{I}, x \in \mathbf{X}, \tag{2.1}$$

where $\beta(\cdot)$ is a nonparametric function of τ, which when $\mathcal{I} = (0, 1)$ also corresponds to the stochastic model with random coefficients:

$$Y = X'\beta(\varepsilon), \qquad \varepsilon \stackrel{d}{=} U(0, 1), X \in \mathbf{X}. \tag{2.2}$$

Here it is necessary that (2.1) holds for

$$\mathcal{I} = [0, \eta] \qquad \text{for some } 0 < \eta \le 1 \text{ and } x \in \mathbf{X}, \text{ a compact subset of } \mathbb{R}^d. \tag{2.3}$$

Different linear models (2.1) can be applied to different covariate regions $\mathbf{X}$ [which can be local neighborhoods of a given x_0, in which case the linear model (2.1) is motivated as a Taylor expansion]. The model (2.1) plays a fundamental role in the theoretical and practical literature on quantile regression mentioned in the Introduction. Its appealing feature is the ability to capture quantile-specific covariate effects in a convenient linear framework.

In the sequel, we combine the linear model (2.1) with the tail restrictions from extreme value theory to develop applicable asymptotic results. It is of vital consequence to impose these restrictions in a manner that preserves the quantile-specific covariate effects, as motivated by the empirical examples listed in the Introduction. For instance, in the analysis of U.S. birthweights, Abrevaya (2001) finds that smoking and the absence of pre-natal care impact the low conditional quantiles of birthweights much more negatively than the central birthweight quantiles. The linear framework (2.1) is able to accommodate this type of impact through the quantile-specific coefficients $\beta(\tau)$, where $\beta_{-1}(\tau)$, for τ near 0, describes the effect of covariate factors on extremely low birthweights and, say, $\beta_{-1}(1/2)$ describes the effect on central birthweights. Thus, when imposing extreme value restrictions, it is important to preserve this ability.

The inference about $\beta(\tau)$ is based on the regression quantile statistics $\hat{\beta}(\tau)$ [Koenker and Bassett (1978)] defined by the least asymmetric absolute deviation problem:

$$\hat{\beta}(\tau) \in \arg\min_{\beta \in \mathbb{R}^d} \sum_{t=1}^{T} \rho_\tau(Y_t - X_t'\beta) \qquad \text{where } \rho_\tau(u) = \big(\tau - 1(u \le 0)\big)u, \tag{2.4}$$

of which Laplace's (1818) median regression is an important case with $\rho_{1/2}(u) = |u|/2$. The statistics $\hat{\beta}(\tau)$ naturally generalize the ordinary sample quantiles to the conditional setting. In fact, the usual univariate τ-quantiles can be recovered as the solution to this problem without covariates, that is, when $X_t = 1$. [E.g., if $\tau T \in (0,1)$, $\hat{\beta}(\tau) = Y_{(1)}$, and if $\tau T \in (1,2)$, $\hat{\beta}(\tau) = Y_{(2)}$, etc.]

In order to provide large sample properties of $\hat{\beta}(\tau)$ in the tails, we distinguish three types of sample regression quantiles, following the classical theory of order statistics: (i) an extreme order sequence, when $\tau_T \searrow 0$, $\tau_T T \to k > 0$, (ii) an intermediate order sequence, when $\tau_T \searrow 0$, $\tau_T T \to \infty$, (iii) a central order sequence, when $\tau \in (0,1)$ is fixed, and $T \to \infty$ (under which the conventional theory applies). We consider $\hat{\beta}(\tau_T)$ under the extreme and intermediate order sequences, and refer to $\hat{\beta}(\tau_T)$ under both sequences as the *extremal regression quantiles*. In what follows, we omit the T in τ_T whenever it does not cause confusion.

3. The extreme value restrictions on the linear quantile regression model. This section joins the linear model (2.1) together with the tail restrictions from extreme value theory, examines the consequences and presents examples.

Consider a random variable u with distribution function F_u and lower end-point $s_u = 0$ or $s_u = -\infty$. Recall [cf. Resnick (1987)] that F_u is said to have tail of type 1, 2 or 3 if for

$$\begin{aligned}
&\text{type 1:} && \text{as } z \searrow s_u = 0 \text{ or } -\infty, \\
& && F_u(z + va(z)) \sim F_u(z)e^{v} && \forall v \in \mathbb{R}, \xi \equiv 0, \\
&\text{type 2:} && \text{as } z \searrow s_u = -\infty, \\
& && F_u(vz) \sim v^{-1/\xi} F_u(z) && \forall v > 0, \xi > 0, \\
&\text{type 3:} && \text{as } z \searrow s_u = 0, \\
& && F_u(vz) \sim v^{-1/\xi} F_u(z) && \forall v > 0, \xi < 0,
\end{aligned} \tag{3.1}$$

where $a(z) \equiv \int_{s_u}^{z} F_u(v)\,dv / F_u(z)$, for $z > s_u$. The number ξ is commonly called the extreme value index, and F_u with tails of types 1–3 is said to belong to the domain of minimum attraction. [$a(z) \sim b(z)$ denotes that $a(z)/b(z) \to 1$ as a specified limit over z is taken.]

CONDITION R1. In addition to (2.1), there exists an auxiliary line $x \mapsto x'\beta_r$ such that for

$$U \equiv Y - X'\beta_r \qquad \text{with } s_U = 0 \text{ or } s_U = -\infty, \tag{3.2}$$

and some F_u with type 1, 2 or 3 tails,

$$F_U(z|x) \sim K(x) \cdot F_u(z) \qquad \text{uniformly in } x \in \mathbf{X}, \text{ as } z \searrow s_U, \tag{3.3}$$

where $K(\cdot) > 0$ is a continuous bounded function on $\mathbf{X}$. Without loss of generality, let $K(x) = 1$ at $x = \mu_X$ and $F_u(z) \equiv F_U(z|\mu_X)$.

CONDITION R2. The distribution function of $X = (1, X_{-1}')'$, F_X, has compact support $\mathbf{X}$ with EXX' positive definite. Without loss of generality, let $\mu_X = EX = (1, 0, \dots, 0)'$.

When Y has a finite lower endpoint, that is, $X'\beta(0) > -\infty$, it is implicit in Condition R1 that $\beta_r \equiv \beta(0)$ so that $U \equiv Y - X'\beta(0) \geq 0$ has endpoint 0 by construction. In the unbounded support case, $X'\beta(0) = -\infty$ and is not suitable as an auxiliary line, but existence of any other line such that Condition R1 holds suffices.

Condition R1 is the main assumption. First, Condition R1 requires the tails of $U = Y - X'\beta_r$ for some β_r to be in the domain of minimum attraction, which is a nonparametric class of distributions [cf. Resnick (1987) and Embrechts, Klüppelberg and Mikosch (1997)]. In this sense, the specification Condition R1 is semiparametric. Examples 3.1 and 3.2 present some of the regression models covered by Condition R1. Second, Condition R1 also requires that, for any $x', x'' \in \mathbf{X}$, $z \mapsto F_U(z|x')$ and $z \mapsto F_U(z|x'')$ are tail equivalent up to a constant. This condition is motivated by the closure of the domain of minimum attraction under tail equivalence [cf. Proposition 1.19 in Resnick (1987)].

Compactness of $\mathbf{X}$ in Condition R1 is necessary, as the limit theory for regression quantiles may generally change otherwise. In applications, compactness may be imposed by the explicit trimming of observations depending on whether $X_t \in \mathbf{X}$. In this case the linear model (2.1) is assumed to apply only to values of X in $\mathbf{X}$. Clearly, the smaller $\mathbf{X}$, the less restrictive is the linear model by virtue of Taylor approximation [e.g., Chaudhuri (1991)]. Also, trimming X to $\mathbf{X}$ eliminates the impact of outlying values on the limit distribution and inference, as it does in the case of the central regression quantiles. In some cases it should be possible to make X unbounded by imposing higher level nonprimitive conditions, for example, similar to those on page 98 in Knight (2001). However, since we view $\mathbf{X}$ as a "small" neighborhood over which the linear approximation (2.1) is adequate, we do not pursue this extension.

Theorem 3.1 shows that the function $K(x)$ in Condition R1 can be represented by the following types. Other properties of the linear quantile regression model under Conditions R1 and R2 are obtained in Lemma 9.1 given in Section 9.1.

THEOREM 3.1 [Three types of $K(\cdot)$]. *Under Conditions* R1 *and* R2, *for some* $\mathbf{c} \in \mathbb{R}^d$,

$$K(x) = \begin{cases} e^{-x'\mathbf{c}}, & \text{when } F_u \text{ has type 1 tails, } \xi = 0, \\ (x'\mathbf{c})^{1/\xi}, & \text{when } F_u \text{ has type 2 tails, } \xi > 0, \\ (x'\mathbf{c})^{1/\xi}, & \text{when } F_u \text{ has type 3 tails, } \xi < 0, \end{cases} \tag{3.4}$$

where $\mu_X'\mathbf{c} = 1$ *for type* 2 *and* 3 *tails*, $\mu_X'\mathbf{c} = 0$ *for type* 1 *tails, and* $x'\mathbf{c} > 0$ *for all* $x \in \mathbf{X}$ *for types* 2 *and* 3.

REMARK 3.1. The condition $X'\mathbf{c} > 0$ a.s. for tails of types 2 and 3 arises from the linearity assumption (2.1). Indeed, (2.1) imposes that the quantiles should not cross: if $l > 1$, then $X'(\beta(l\tau) - \beta(\tau)) > 0$ a.s. Since by Lemma 9.1(v) $X'(\beta(l\tau) - \beta(\tau))/\mu_X'(\beta(l\tau) - \beta(\tau)) \to X'\mathbf{c}$ as $\tau \searrow 0$, the noncrossing condition requires $X'\mathbf{c} > 0$ a.s. In location-scale shift models (cf. Example 3.2), the condition $X'\mathbf{c} > 0$ a.s. is equivalent to a logical restriction on the scale function ($X'\sigma > 0$ a.s.). In location shift models (cf. Example 3.1), this condition is ordinarily satisfied since $X'\mathbf{c} = 1$ a.s. for tails of types 2 and 3.

REMARK 3.2. The general case when $P\{K(X) \neq 1\} > 0$ will be referred to as the heterogeneous case, and **c** will be referred to as the *heterogeneity index*. The special case with

$$K(X) = 1 \qquad \text{a.s.} \tag{3.5}$$

will be referred to as the *homogeneous* case. The latter amounts to $\mathbf{c} = \mathbf{0}$ for type 1 tails, and $\mathbf{c} = \mathbf{e}_1' \equiv (1, 0, \ldots)'$ for type 2 and 3 tails. Notice that in this case $X'\mathbf{c} = 1$ a.s. for types 2 and 3 and $X'\mathbf{c} = 0$ a.s. for type 1 tails.

In developing regularity conditions which target regression applications, it is natural to try to cover the most conventional regression settings and, hopefully, more general stochastic specifications. The following examples clarify this possibility.

EXAMPLE 3.1 (*Location shift regression*). Consider the location-shift model

$$Y = X'\beta + U, \tag{3.6}$$

where U is independent of X, and suppose U is in the domain of minimum attraction. When the lower endpoint of the support of U is finite, it is normalized to 0. Clearly, this is a special case of Condition R1 where $X'\beta_r \equiv X'\beta, U \equiv Y - X'\beta, K(X) = 1$ a.s. The data generating process (3.6) has been widely adopted in regression work at least since Huber (1973) and Rao (1965). A variety of standard survival and duration models also imply (3.6) after a transformation, for example, the Cox models with Weibull hazards and accelerated failure time models [cf. Doksum and Gasko (1990)]. Also, (3.6) underlies many theoretical studies of quantile regression. Hence, it is useful that Condition R1 covers (3.6).

EXAMPLE 3.2 (*Location-scale shift regression*). As a generalization of (3.6), consider the stochastic equation

$$Y = X'\beta + X'\sigma \cdot V, \qquad V \text{ is independent of } X, \tag{3.7}$$

where $X'\sigma > 0$ (a.s.) is the scale function, and V is in the domain of minimum attraction with $\xi \neq 0$. (3.7) implies the following linear conditional quantile function

$$F_Y^{-1}(\tau|X) = X'\beta + X'\sigma \cdot F_V^{-1}(\tau). \tag{3.8}$$

Then for $X'\beta_r \equiv X'\beta, U \equiv Y - X'\beta_r = X'\sigma \cdot V$, we have $P(X'\sigma \cdot V \leq z|X) \sim (X'\sigma)^{1/\xi} \cdot F_V(z)$ as $z \searrow 0$ or $-\infty$, so Condition R1 is satisfied with $F_u \equiv F_V$ and $K(X) = (X'\sigma)^{1/\xi}$. The data generating process (3.7) has been adopted in, for example, Koenker and Bassett (1982), Gutenbrunner and Jurečková (1992) and He (1997).

EXAMPLE 3.3 (*Quantile-shift regression*). To see that Condition R1 covers more general stochastic models than (3.6) and (3.7), note that Condition R1 requires that $F_U(u|X)$ or $F_V(u|X)$ be independent of X only in the tails. In both cases, these weaker independence requirements allow X, for example, to have a negative impact on the high and low quantiles but to have a positive impact on the median quantiles. In contrast, notice from (3.8) that (3.6) and (3.7) preclude such quantile-specific impacts. Thus, Condition R1 preserves the heterogeneous impact property of (2.1), allowing the impact of covariate factors on extreme quantiles to be very different from their impact on the central quantiles.

4. Asymptotics of extreme order regression quantiles. Consider sequences $\tau_i, i = 1, \ldots, l$, such that $\tau_i T \to k_i > 0$ as $T \to \infty$, and the corresponding normalized regression quantile statistics $\widehat{Z}_T(k_i)$, where

$$\widehat{Z}_T(k) \equiv a_T\big(\hat{\beta}(\tau) - \beta_r - b_T\mathbf{e}_1\big), \tag{4.1}$$

$\hat{\beta}(\tau)$ is the regression quantile, β_r is the coefficient of the auxiliary line defined in (3.2), $\mathbf{e}_1 \equiv (1, 0, \ldots)' \in \mathbb{R}^d$, and (a_T, b_T) are the canonical normalization constants, given by

$$\begin{aligned} &\text{for type 1 tails:} \quad a_T = 1/a\left[F_u^{-1}\left(\frac{1}{T}\right)\right], && b_T = F_u^{-1}\left(\frac{1}{T}\right), \\ &\text{for type 2 tails:} \quad a_T = -1/F_u^{-1}\left(\frac{1}{T}\right), && b_T = 0, \\ &\text{for type 3 tails:} \quad a_T = 1/F_u^{-1}\left(\frac{1}{T}\right), && b_T = 0, \end{aligned} \tag{4.2}$$

where F_u is defined in Condition R1. Moreover, consider the centered statistic

$$\widehat{Z}_T^c(k) \equiv a_T\big(\hat{\beta}(\tau) - \beta(\tau)\big) \tag{4.3}$$

and the point process, for $U_t = Y_t - X_t'\beta_r$,

$$\widehat{\mathbf{N}}(\cdot) = \sum_{t=1}^{T} \mathbb{1}\big(\{a_T(U_t - b_T), X_t\} \in \cdot\big). \tag{4.4}$$

We will show that $\widehat{\mathbf{N}}(\cdot)$ converges weakly to the Poisson process

$$\mathbf{N}(\cdot) = \sum_{i=1}^{\infty} \mathbb{1}(\{J_i, \mathcal{X}_i\} \in \cdot), \tag{4.5}$$

with points $\{J_i, \mathcal{X}_i\}$ satisfying

$$(J_i, \mathcal{X}_i, i \geq 1) = \begin{cases} \big(\ln(\Gamma_i) + \mathcal{X}_i'\mathbf{c}, \mathcal{X}_i\big), & \text{for type 1 tails,} \\ (-\Gamma_i^{-\xi}\mathcal{X}_i'\mathbf{c}, \mathcal{X}_i), & \text{for type 2 tails,} \\ (\Gamma_i^{-\xi}\mathcal{X}_i'\mathbf{c}, \mathcal{X}_i), & \text{for type 3 tails,} \end{cases} \quad i \geq 1, \tag{4.6}$$

where $\{\mathcal{X}_i\}$ is an i.i.d. sequence with law F_X,

$$\Gamma_i \equiv \sum_{j=1}^{i} \mathcal{E}_j, \qquad i \geq 1, \tag{4.7}$$

and $\{\mathcal{E}_j\}$ is an i.i.d. sequence of unit-exponential variables, independent of $\{\mathcal{X}_i\}$. In the homogeneous case (3.5), J_i and $\mathcal{X}_i$ are independent since

$$\mathcal{X}_i'\mathbf{c} = \begin{cases} 0, & \text{for type 1 tails,} \\ 1, & \text{for type 2 and 3 tails,} \end{cases} \qquad \text{for all } i \geq 1. \tag{4.8}$$

The following theorem establishes the weak limit of $\widehat{Z}_T(k)$'s as a function of $\mathbf{N}$.

THEOREM 4.1 (Extreme order regression quantiles). *Assume Conditions* R1 *and* R2 *and that* $\{Y_t, X_t\}$ *is an i.i.d. or a stationary sequence satisfying the Meyer type conditions of Lemma* 9.4. *Then as* $\tau T \to k > 0$ *and* $T \to \infty$,

$$\widehat{Z}_T(k) \xrightarrow{d} Z_\infty(k) \equiv \arg\min_{z \in \mathcal{Z}} \left[-k\mu_X' z + \int (x'z - u)^+ \, d\mathbf{N}(u, x) \right], \tag{4.9}$$

provided $Z_\infty(k)$ *is a uniquely defined random vector in* $\mathcal{Z}$, *where* $(x'z - u)^+ = \mathbb{1}(u \leq x'z)(x'z - u)$, $\mathcal{Z} = \mathbb{R}^d$ *for type* 1 *and* 3 *tails, and* $\mathcal{Z} = \{z \in \mathbb{R}^d : \max_{x \in \mathbf{X}} z'x \leq 0\}$ *for type* 2 *tails. Moreover,*

$$\widehat{Z}_T^c(k) \xrightarrow{d} Z_\infty^c(k) \equiv Z_\infty(k) - \eta(k), \tag{4.10}$$

where

$$\eta(k) = \begin{cases} \mathbf{c} + \ln k \mathbf{e}_1, & \textit{for type } 1 \textit{ tails,} \\ -k^{-\xi}\mathbf{c}, & \textit{for type } 2 \textit{ tails,} \\ k^{-\xi}\mathbf{c}, & \textit{for type } 3 \textit{ tails.} \end{cases} \tag{4.11}$$

If $Z_\infty(k)$ *is a uniquely defined random vector for* $k = k_1, \dots, k_l$,

$$\big(\widehat{Z}_T(k_1)', \dots, \widehat{Z}_T(k_l)'\big)' \xrightarrow{d} \big(Z_\infty(k_1)', \dots, Z_\infty(k_l)'\big)',$$

$$\big(\widehat{Z}_T^c(k_1)', \dots, \widehat{Z}_T^c(k_l)'\big)' \xrightarrow{d} \big(Z_\infty^c(k_1)', \dots, Z_\infty^c(k_l)'\big)'.$$

REMARK 4.1 (*The limit criterion function*). The limit objective function $-k\mu_X' z + \int (x'z - u)^+ \, d\mathbf{N}(u, x)$ can also be written as

$$-k\mu_X' z + \sum_{i=1}^{\infty} (\mathcal{X}_i' z - J_i)^+. \tag{4.12}$$

REMARK 4.2 (*Homogeneous case*). The limit result is simpler for the homogeneous case (3.5), since $\mathbf{N}$ does not depend on the heterogeneity parameter $\mathbf{c}$ due to (4.8).

REMARK 4.3 (*Case with $\tau T \to 0$*). The linear programming estimator, which corresponds to $T\tau \to 0$ in (2.4) (in comparison, here $\tau T \to k > 0$), was studied in Feigin and Resnick (1994), Smith (1994), Portnoy and Jurečková (1999), Knight (1999, 2001) and Chernozhukov (1998) under various types of location-shift specification (3.6). This estimator is the solution to the problem

$$\max_{\beta \in \mathbb{R}^d} \overline{X}' \beta \text{ such that } Y_t \geq X_t' \beta \qquad \text{for all } t \leq T, \overline{X} = T^{-1} \sum_{t=1}^{T} X_t. \tag{4.13}$$

The asymptotics of (4.13) and proofs differ substantively from the ones given here for $\tau T \to k > 0$. The analysis of $\tau T \to k > 0$ is specifically motivated by the applications listed in the Introduction.

REMARK 4.4 (*Uniqueness*). The limit objective function is convex, and it is assumed in Theorem 4.1 that $Z_\infty(k)$ is unique and tight. Lemma 9.7 shows that a sufficient condition for tightness is the design condition of Portnoy and Jurečková (1999). Taking tightness as given, conditions for uniqueness can be established. Define $\mathcal{H}$ as the set of all d-element subsets of $\mathbb{N}$. 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 nonempty a.s. by Condition R2 and is countable. Application of the argument of Theorem 3.1 of Koenker and Bassett (1978) gives that an arg min of (4.12) takes the form $z_h = \mathcal{X}(h)^{-1} J(h)$ for some $h \in \mathcal{H}^*$, and must satisfy the gradient condition

$$\zeta_k(z_h) \equiv \left(k\mu_X - \sum_{t=1}^{\infty} \mathbb{1}(J_t < \mathcal{X}_t' z_h) \mathcal{X}_t \right)' \mathcal{X}(h)^{-1} \in [0, 1]^d, \tag{4.14}$$

where the arg min is unique iff $\zeta_k(z_h) \in \mathcal{D} = (0, 1)^d$. Thus, uniqueness holds for a fixed $k > 0$ if

$$P\big(\zeta_k(z_h) \in \partial \mathcal{D} \text{ for some } h \in \mathcal{H}^*\big) = 0. \tag{4.15}$$

This condition is a direct analog of Koenker and Bassett's (1978) condition for uniqueness in finite samples; for instance, it is satisfied for a given k when covariates $\mathcal{X}_{-1t}$ are absolutely continuous [cf. Portnoy (1991b)]. Thus, uniqueness holds generically in the sense that for a fixed k adding arbitrarily small absolutely continuous perturbations to $\{\mathcal{X}_{-1t}\}$ ensures (4.15).

REMARK 4.5 (*Asymptotic density*). The density of $Z_\infty(k)$ can be stated following Koenker and Bassett (1978). Given $\{\mathcal{X}_t\}$, $h \in \mathcal{H}^*$, and $J(h)$, the probability that $Z_\infty(k) = \mathcal{X}(h)^{-1}J(h)$ equals $P\{\zeta_k(\mathcal{X}(h)^{-1}J(h)) \in \mathcal{D}|\{\mathcal{X}_t\}, J(h)\}$. Conditional on $\{Z_\infty(k) = \mathcal{X}(h)^{-1}J(h)\}$, $h \in \mathcal{H}^*$, and $\mathcal{X}(h)$, the density of $Z_\infty(k)$ at z is $f_{J(h)|\mathcal{X}(h)}(\mathcal{X}(h)z) \cdot |\mathcal{X}(h)|$, where $f_{J(h)|\mathcal{X}(h)}(u)$, $u \in \mathbb{R}^d$, is the joint density of $J(h)$ conditional on $\mathcal{X}(h)$. Thus, the joint density of $Z_\infty(k)$ at z is

$$f_{Z_\infty(k)}(z) = E\Bigg[\sum_{h\in\mathcal{H}^*} f_{J(h)|\mathcal{X}(h)}(\mathcal{X}(h)z) \cdot |\mathcal{X}(h)| \\ \times P\big\{\zeta_k\big(\mathcal{X}(h)^{-1}J(h)\big) \in \mathcal{D}|\{\mathcal{X}_t\}, J(h)\big\}\Bigg].$$

Finally, for $f_{Z_\infty(k)}(z)$ to be nondefective, $Z_\infty(k) = O_p(1)$ should be established (cf. Lemma 9.7).

REMARK 4.6 (*Univariate case*). The density simplifies in the classical nonregression case, that is, when $X = 1$, in which case we also have the simplification (4.8). In this case, an arg min is necessarily an order statistic, that is, $z_h = J(h) = J_h$; the gradient condition (4.14) becomes

$$\zeta_k(z_h) \equiv \left(k - \sum_{t=1}^{\infty} \mathbb{1}(J_t < z_h)\right) \in [0, 1]; \tag{4.16}$$

and the condition for uniqueness is that $\zeta_k(z_h) \in \mathcal{D} = (0, 1)$. Then, for $k \neq \lceil k \rceil$, $P\{\zeta_k(z_h) \in \mathcal{D}\} = 1$ if $h = \lceil k \rceil$ and $P\{\zeta_k(z_h) \in \mathcal{D}\} = 0$ if $h \neq \lceil k \rceil$. Here $k \neq \lceil k \rceil$ is needed for uniqueness. Hence, $f_{Z_\infty(k)}(z) = f_{J_{\lceil k \rceil}}(z)$, which is the limit density of the $\lceil k \rceil$th order statistics in the univariate case. Thus, uniqueness holds for almost every $k \in (0, \infty)$.

5. Asymptotics of intermediate order regression quantiles. In order to develop asymptotic results for the intermediate regression quantiles, the following additional Condition R3 will be added. First, existence of the quantile density function $\partial F_U^{-1}(\tau|x)/\partial\tau \equiv x'\,\partial\beta(\tau)/\partial\tau$ and its regular variation will be required. Second, the tail equivalence of the conditional distribution functions, previously assumed in Condition R1, will now be strengthened to the tail equivalence of conditional quantile density functions.

CONDITION R3. In addition to Conditions R1 and R2, for ξ defined in (3.1),

$$\text{(5.1)}\qquad \begin{aligned} &\text{(i)}\ \frac{\partial F_U^{-1}(\tau|x)}{\partial\tau} \sim \frac{\partial F_u^{-1}(\tau/K(x))}{\partial\tau} \qquad \text{uniformly in } x\in\mathbf{X},\\ &\text{(ii)}\ \frac{\partial F_u^{-1}(\tau)}{\partial\tau} \qquad \text{is regularly varying at 0 with exponent } -\xi-1. \end{aligned}$$

In the homoscedastic case (3.5), Condition R3(i) amounts to $\frac{\partial F_U^{-1}(\tau|x)}{\partial\tau} \sim \frac{\partial F_u^{-1}(\tau)}{\partial\tau}$ uniformly in $x\in\mathbf{X}$ as $\tau\searrow 0$. Condition R3(ii) is a von Mises type condition; see Dekkers and de Haan (1989) for a detailed analysis of the plausibility of Condition R3(ii).

For an intermediate sequence such that $\tau\searrow 0$ and $\tau T\to\infty$, define, for $m>1$,

$$\text{(5.2)}\qquad \widehat{Z}_T \equiv a_T\big(\hat\beta(\tau)-\beta(\tau)\big), \qquad a_T \equiv \frac{\sqrt{\tau T}}{\mu_X'(\beta(m\tau)-\beta(\tau))}.$$

Consider also k sequences $\{\tau l_1,\dots,\tau l_k\}$, where $l_1,\dots,l_k$ are positive constants, and corresponding statistics $(\widehat{Z}_T(l_1)',\dots,\widehat{Z}_T(l_k)')'$, where, for $l>0$ and $m>1$,

$$\text{(5.3)}\qquad \widehat{Z}_T(l) \equiv a_T(l)\big(\hat\beta(l\tau)-\beta(l\tau)\big), \qquad a_T(l) \equiv \frac{\sqrt{\tau l T}}{\mu_X'(\beta(ml\tau)-\beta(l\tau))}.$$

The following theorem establishes the weak limits for $\widehat{Z}_T$ and $\widehat{Z}_T(l)$'s. Because $\tau\searrow 0$, the limits depend only on the tail parameters ξ and $\mathbf{c}$, as in Theorem 4.1, but since $\tau T\to\infty$, the limits are normal, unlike in Theorem 4.1.

THEOREM 5.1 (Intermediate order regression quantiles). *Suppose Conditions* R1–R3 *hold, and that* $\{Y_t, X_t\}$ *is an i.i.d. sequence or a stationary series satisfying the conditions of Lemma* 9.6. *Then, as* $\tau T\to\infty$ *and* $\tau\searrow 0$,

$$\text{(5.4)}\qquad \widehat{Z}_T \xrightarrow{d} Z_\infty = N(0,\Omega_0), \qquad \Omega_0 \equiv \mathcal{Q}_H^{-1}\mathcal{Q}_X\mathcal{Q}_H^{-1}\frac{\xi^2}{(m^{-\xi}-1)^2},$$

where, for $\xi=0$, *interpret* $\xi^2/(m^{-\xi}-1)^2$ *as* $(\ln m)^{-2}$ *and*

$$\text{(5.5)}\qquad \mathcal{Q}_H \equiv E[H(X)]^{-1}XX', \qquad \mathcal{Q}_X \equiv EXX',$$

$$\text{(5.6)}\qquad H(x)\equiv x'\mathbf{c} \quad \textit{for type 2 and 3 tails}, \qquad H(x)\equiv 1 \quad \textit{for type 1 tails}.$$

In addition,

$$\text{(5.7)}\qquad \big(\widehat{Z}_T(l_1)',\dots,\widehat{Z}_T(l_k)'\big)' \xrightarrow{d} \big(Z_\infty(l_1)',\dots,Z_\infty(l_k)'\big)' = N(0,\Omega),$$

$$\text{(5.8)}\qquad EZ_\infty(l_i)Z_\infty(l_j)' = \Omega_0\times\min(l_i,l_j)/\sqrt{l_i l_j}.$$

Finally, $a_T(l)$ *can be replaced by* $\sqrt{\tau l T}/\overline{X}'(\hat{\beta}(ml\tau) - \hat{\beta}(l\tau))$ *without affecting* (5.4) *and* (5.7), *that is,*

$$a_T(l)\Big/\left(\frac{\sqrt{\tau l T}}{\overline{X}'(\hat{\beta}(ml\tau) - \hat{\beta}(l\tau))}\right) \xrightarrow{p} 1 \qquad \text{where } \overline{X} = T^{-1}\sum_{t=1}^{T} X_t. \tag{5.9}$$

REMARK 5.1 (*Scaling constants*). It may be useful to have the same normalization a_T in place of $a_T(l)$ for the joint convergence. This is possible by noting that $a_T/a_T(l) \to l^{-\xi}/\sqrt{l}$.

REMARK 5.2 (*Homogeneous case*). In the homogeneous case (3.5), $H(X) = 1$, so the variance simplifies to

$$\Omega_0 = \mathcal{Q}_X^{-1}\frac{\xi^2}{(m^{-\xi} - 1)^2}. \tag{5.10}$$

REMARK 5.3 (*Nonregression case*). Theorem 5.1 extends Theorem 3.1 of Dekkers and de Haan (1989), which applies to univariate quantiles, to the case of regression quantiles. In fact, Theorem 3.1 of Dekkers and de Haan (1989) can be specialized from Theorem 5.1 with $X = 1$ and $m = 2$. In this case the variance becomes

$$\frac{\xi^2}{(2^{-\xi} - 1)^2} = \frac{2^{2\xi}\xi^2}{(2^{\xi} - 1)^2}, \tag{5.11}$$

as Dekkers and de Haan (1989) found in their Theorem 3.1.

6. Quantile regression spacings and tail inference. The tail parameters enter the limit distributions in Theorems 4.1 and 5.1, and estimation of the tail index is an important problem of its own. The following results show how to estimate them by applying Pickands (1975) type procedures to the quantile regression spacings.

Consider the following parameters and statistics:

$$\begin{aligned} \varphi &= \frac{x'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))}{x'(\beta(m\tau) - \beta(\tau))}, \\ \rho_{x,\dot{x},l} &= \frac{x'(\beta(ml\tau) - \beta(l\tau))}{\dot{x}'(\beta(m\tau) - \beta(\tau))}, \\ \hat{\rho}_{x,\dot{x},l} &= \frac{x'(\hat{\beta}(ml\tau) - \hat{\beta}(l\tau))}{\dot{x}'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))}. \end{aligned} \tag{6.1}$$

Theorem 6.1 shows that the quantile regression spacings of intermediate order consistently approximate the corresponding spacings in the population [results (i) and (ii)], which then reveal the tail parameters [results (iii) and (iv)].

THEOREM 6.1 (Quantile regression spacings and tail inference). *Suppose the conditions of Theorem* 5.1 *hold. Then as* $\tau \searrow 0$, $\tau T \to \infty$, *for all* $l > 0$, $m > 1$, $x, \dot{x} \in \mathbf{X}$,

(i) $\varphi \xrightarrow{p} 1$,

(ii) $\hat{\rho}_{x,\dot{x},l} - \rho_{x,\dot{x},l} \xrightarrow{p} 0$, $\rho_{x,\dot{x},l} \to l^{-\xi} \cdot [H(x)/H(\dot{x})]$, *for* $H(x)$ *defined in Theorem* 5.1,

(iii) $\hat{\xi}_{rp} \equiv \frac{-1}{\ln l} \ln \hat{\rho}_{\overline{X},\overline{X},l} \xrightarrow{p} \xi$,

(iv) $\hat{\rho}_{x,\overline{X},1} \xrightarrow{p} x'\mathbf{c}$ *uniformly in* $x \in \mathbf{X}$ $(\xi \neq 0)$,

(v) *for* $\pi = \mu_X' \mathcal{Q}_H^{-1} \mathcal{Q}_X \mathcal{Q}_H^{-1} \mu_X$, $l = m = 2$, *if* $\sqrt{\tau T}(\rho_{\overline{X},\overline{X},l} - \lim_T \rho_{\overline{X},\overline{X},l}) \to 0$,

$$\sqrt{\tau T}(\hat{\xi}_{rp} - \xi) \xrightarrow{d} N\left(0, \pi \cdot \frac{\xi^2(2^{2\xi+1}+1)}{(2(2^\xi - 1)\ln 2)^2}\right). \tag{6.2}$$

REMARK 6.1 (*Homogeneous case*). The proposed estimator $\hat{\xi}_{rp}$ consistently estimates the tail index ξ in the heteroscedastic and homoscedastic quantile regression models, and it is a regression extension of the Pickands (1975) estimator. In fact, in the homoscedastic model (3.5) or when $X = 1$, $\pi = \mu_X'(EXX')^{-1}\mu_X = \mathbf{e}_1'(EXX')^{-1}\mathbf{e}_1 = 1$, so the variance in (6.2) reduces to that of the canonical Pickands estimator.

7. An illustrative example. The set of results established here may provide reliable and practical inference for extremal regression quantiles. To illustrate this possibility, the following simple example compares graphically the conventional central asymptotic approximation, where, for fixed $\tau \in (0, 1)$ as $T \to \infty$,

$$\sqrt{T}\big(\hat{\beta}(\tau) - \beta(\tau)\big) \xrightarrow{d} N\left(0, \frac{1}{f_U^2(F_U^{-1}(\tau))}(EXX')^{-1}\tau(1-\tau)\right), \tag{7.1}$$

to the extreme approximation (cf. Theorem 4.1). The comparison is based on the following design: $\tau = 0.025$, $Y_t = X_t'\beta + U_t, U_t \sim$ Cauchy , $t = 1, \dots, 500$, where $X_t = (1, X_{-1t}')' \in \mathbb{R}^5$, X_{-1t} are i.i.d. Beta(3, 3) variables, and $\beta = (1, 1, 1, 1, 1)$. [A more detailed simulation study is given in Chernozhukov (1999).] In this comparison, the parameters of the limit distribution are fixed at the true values.

Figure 1 plots (a) quantiles of the simulated finite-sample distribution of $\hat{\beta}_1(0.025)$ and $\hat{\beta}_2(0.025)$, (b) quantiles of the simulated extreme approximation (cf. Theorem 4.1), (c) quantiles of the central approximation [cf. (7.1)]. Here $\tau \times T = 0.025 \times 500 = 12.5$. It can be seen that the extreme approximation accurately captures the actual sampling distribution of both the intercept estimator $\hat{\beta}_1(0.025)$ and the slope estimator $\hat{\beta}_2(0.025)$. In contrast, the central approximation (7.1) does not capture asymmetry and thick tails of the true finite sample distribution. The

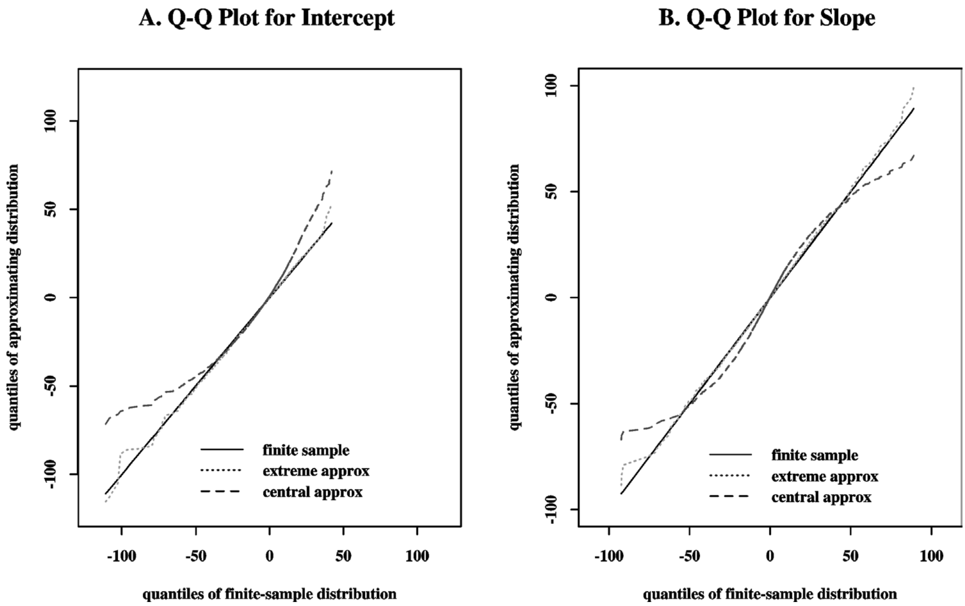

FIG. 1. *Panel* A *plots quantiles of the finite-sample distribution of* $\widehat{\beta}_1(\tau)$ (*horizontal axis*) *against the quantiles of the extreme approximation* (*cf. Theorem* 4.1) *and the quantiles of the central approximation* (7.1) (*vertical axis*). *Panel* B *plots quantiles of the finite-sample distribution of* $\widehat{\beta}_2(\tau)$ (*horizontal axis*) *against the quantiles of the extreme approximation* (*cf. Theorem* 4.1) *and the quantiles of the central approximation* (7.1) (*vertical axis*). *The plot is based on* 10,000 *simulations of the regression model described in Section* 7. *The dashed line* "- - - -" *denotes quantiles of the central approximation, and the dotted line* "$\cdots\cdots$" *denotes quantiles of the extreme approximation* (*this approximation almost coincides with* "——"). *The simulated quantiles of the finite-sample distribution are given by the* 45-*degree line depicted as the solid line* "——."

intermediate approximation (cf. Theorem 5.1), performs similarly to the central approximation and is not plotted. The central and intermediate approximations are expected to perform better for less extreme quantiles.

8. Conclusion. The paper obtains the large sample properties of extreme order and intermediate order quantile regression for the class of linear quantile regression models with tails of the response variable restricted to the domain of minimum attraction and closed under tail equivalence across conditioning values. There are several interesting directions for future work. It would be important to determine the most practical and reliable inference procedures that can be based on the obtained limit distributions. Also, it would be interesting to examine estimation of the extreme conditional quantiles defined through an extrapolation of the intermediate regression quantiles. The nonregression case has been considered in Dekkers and de Haan (1989) and de Haan and Rootzén (1993), and the approach may prove useful in the quantile regression case. Another interesting direction would be an investigation of the Hill and other tail index estimators based on regression quantiles.

9. Proofs.

9.1. *Properties of the linear quantile regression model under Conditions* R1 *and* R2. Let

$$M \equiv \text{any fixed compact sub-interval of } (0,1)\cup(1,\infty), \tag{9.1}$$

$$M' \equiv \text{any other fixed compact sub-interval of } (0,1)\cup(1,\infty), \tag{9.2}$$

$$\mathcal{T}(\tau') \equiv \{\tau : \tau = s\tau', s \in \mathcal{L}\} \qquad \text{where } \tau' \searrow 0, \tag{9.3}$$

$$\mathcal{L} \equiv \text{any fixed compact sub-interval of } (0,\infty). \tag{9.4}$$

LEMMA 9.1 (Properties of the linear model under Conditions R1 and R2). *Conditions* R1 *and* R2 *imply that* (*for a constant vector* $\mathbf{c}$ *specified in Theorem* 3.1):

(i) $K(x)$ *can be represented by the forms specified in Theorem* 3.1.
(ii) $a_T(\beta(\tau) - \beta_r - b_T\mathbf{e}_1) \to \eta(k)$ *for* $\eta(k)$ *defined in Theorem* 4.1.
(iii) *Uniformly in* $(m,\tau,x) \in M \times \mathcal{T}(\tau') \times \mathbf{X}$, *as* $\tau' \searrow 0$,

$$\frac{\beta_{-1}(\tau) - \beta_{-1r}}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \to \mu(m) = \begin{cases} \dfrac{\mathbf{c}_{-1}}{m^{-\xi}-1}, & \textit{for } \xi < 0,\\[2mm] \dfrac{-\mathbf{c}_{-1}}{m^{-\xi}-1}, & \textit{for } \xi > 0,\\[2mm] \dfrac{\mathbf{c}_{-1}}{\ln m}, & \textit{for } \xi = 0; \end{cases} \tag{9.5}$$

also $\beta_1(\tau) - \beta_{1r} = F_u^{-1}(\tau)$, *and* $(\beta_{-1}(\tau) - \beta_{-1r})/F_u^{-1}(\tau) \to \mathbf{c}_{-1}$ *for* $\xi \neq 0$.
(iv) *Uniformly in* $(m,\tau,x) \in M \times \mathcal{T}(\tau') \times \mathbf{X}$, *as* $\tau' \searrow 0$,

$$\frac{(x-\mu_X)'(\beta(\tau)-\beta_r)}{\mu_X'(\beta(m\tau)-\beta(\tau))} \to \begin{cases} (x-\mu_X)'\dfrac{\mathbf{c}}{m^{-\xi}-1}, & \textit{if } \xi < 0,\\[2mm] (x-\mu_X)'\dfrac{-\mathbf{c}}{m^{-\xi}-1}, & \textit{if } \xi > 0,\\[2mm] (x-\mu_X)'\dfrac{\mathbf{c}}{\ln m}, & \textit{if } \xi = 0. \end{cases} \tag{9.6}$$

(v) *Uniformly in* $(m,\tau,x) \in M \times \mathcal{T}(\tau') \times \mathbf{X}$, *as* $\tau' \searrow 0$,

$$\frac{x'(\beta(m\tau)-\beta(\tau))}{\mu_X'(\beta(m\tau)-\beta(\tau))} \to \begin{cases} x'\mathbf{c}, & \textit{if } \xi < 0,\\ x'\mathbf{c}, & \textit{if } \xi > 0,\\ 1, & \textit{if } \xi = 0. \end{cases} \tag{9.7}$$

(vi) *Uniformly in* $(l, m, \tau, x) \in M \times M' \times \mathcal{T}(\tau') \times \mathbf{X}$, *as* $\tau' \searrow 0$,

$$\frac{x'(\beta(l\tau) - \beta(\tau))}{x'(\beta(m\tau) - \beta(\tau))} \to \begin{cases} \dfrac{l^{-\xi} - 1}{m^{-\xi} - 1}, & \text{if } \xi < 0, \\ \dfrac{1 - l^{-\xi}}{1 - m^{-\xi}}, & \text{if } \xi > 0, \\ \dfrac{\ln l}{\ln m}, & \text{if } \xi = 0. \end{cases} \tag{9.8}$$

Write $F_u \in D(H_\xi)$ if F_u is a c.d.f. in the domain of minimum attraction with tail index ξ. Write $F_u \in \mathcal{R}_\gamma(0)$ if F_u is a regularly varying function at 0 with exponent γ.

LEMMA 9.2 (Useful relations). *Under Conditions* R1 *and* R2, *uniformly in* $(m, l, \tau) \in M \times M' \times \mathcal{T}(\tau')$, *as* $\tau' \searrow 0$:

(i) *Suppose* $F_1(z) \sim F_2(z)$ *as* $z \searrow 0$ *or* $-\infty$ *and* $F_1 \in D(H_\xi)$. *Then* $F_2 \in D(H_\xi)$; F_1^{-1} *and* $F_2^{-1} \in \mathcal{R}_{-\xi}(0)$; $F_1(F_1^{-1}(\tau)) \sim \tau$ *and* $F_2(F_2^{-1}(\tau)) \sim \tau$; *and*

$$\big(F_1^{-1}(m\tau) - F_1^{-1}(\tau)\big) \sim \big(F_2^{-1}(m\tau) - F_2^{-1}(\tau)\big). \tag{9.9}$$

(ii) *If* $F_U(z|x) \sim K(x)F_u(z)$ *as* $z \searrow 0$ *or* $-\infty$ *for each* $x \in \mathbf{X}$ (*compact*), *where* $K(x) \in (0, \infty)$ *for all* $x \in \mathbf{X}$, *then for each* $x \in \mathbf{X}$,

$$F_U^{-1}(m\tau|x) - F_U^{-1}(\tau|x) \sim F_u^{-1}(m\tau/K(x)) - F_u^{-1}(\tau/K(x)). \tag{9.10}$$

(iii) $\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{F_u^{-1}(l\tau) - F_u^{-1}(\tau)} \to \frac{m^{-\xi}-1}{l^{-\xi}-1}$ *if* $\xi < 0$, $\frac{1-m^{-\xi}}{1-l^{-\xi}}$ *if* $\xi > 0$, $\frac{\ln m}{\ln l}$ *if* $\xi = 0$; *for* $F_u \in D(H_\xi)$.

(iv) $\frac{F_u^{-1}(lm\tau) - F_u^{-1}(l\tau)}{a(F_u^{-1}(\tau))} \to \ln m$ *if* $F_u \in D(H_0)$, *where* $a(\cdot)$ *is the auxiliary function defined in* (3.1).

PROOF. Results (i), (iii) and (iv) are well known [cf. de Haan (1984) and Resnick (1987), Chapters 1 and 2]. Result (ii) holds from (i) pointwise in x. $\square$

PROOF OF LEMMA 9.1. Claim (i): The proof consists of two steps, where we use notation $(\mathcal{L}, M, \mathcal{T}(\tau'), \tau')$ as defined in (9.1)–(9.4).

STEP 1. In this step all of the results hold uniformly in $(m, \tau, x) \in M \times \mathcal{T}(\tau') \times \mathbf{X}$ as $\tau' \searrow 0$, but we shall suppress this qualification for notational simplicity. By construction in Condition R1, $x'(\beta(\tau) - \beta_r) \equiv F_U^{-1}(\tau|x)$ and $\mu_X'(\beta(\tau) - \beta_r) \equiv F_U^{-1}(\tau|\mu_X) \equiv F_u^{-1}(\tau)$. Hence,

$$B_\tau(x, m) \equiv \frac{(x - \mu_X)'(\beta(\tau) - \beta_r)}{\mu_X'(\beta(m\tau) - \beta(\tau))} \equiv \frac{F_U^{-1}(\tau|x) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}. \tag{9.11}$$

We would like to show that, for each $x \in \mathbf{X}$,

$$B_\tau(x,m) \to B(x,m) \equiv \begin{cases} \dfrac{(1/K(x))^{-\xi} - 1}{m^{-\xi} - 1}, & \text{if } \xi < 0, \\ \dfrac{1 - (1/K(x))^{-\xi}}{1 - m^{-\xi}}, & \text{if } \xi > 0, \\ \dfrac{\ln(1/K(x))}{\ln m}, & \text{if } \xi = 0. \end{cases} \tag{9.12}$$

We will show (9.12) for the case $\xi < 0$ only; others follow similarly. Fix any $x \in \mathbf{X}$. By Condition R1 and Lemma 9.2(i), $F_U(F_U^{-1}(\tau|x)|x) \sim \tau$. Hence, by Condition R1, $K(x) \cdot F_u(F_U^{-1}(\tau|x)) \sim \tau$ as $\tau' \searrow 0$. Therefore, there exist sequences of constants $K_\tau(x)$ and $K'_\tau(x)$ such that

$$F_u^{-1}\big(\tau/K_\tau(x)\big) \le F_U^{-1}(\tau|x) \le F_u^{-1}\big(\tau/K'_\tau(x)\big) \quad \text{where } K_\tau(x) \to K(x) \text{ and } K'_\tau(x) \to K(x). \tag{9.13}$$

Therefore,

$$\frac{F_u^{-1}(\tau/K_\tau(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \le B_\tau(x,m) \le \frac{F_u^{-1}(\tau/K'_\tau(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}. \tag{9.14}$$

Suppose that $K(x) \neq 1$. By Lemma 9.2(iii),

$$\frac{F_u^{-1}(\tau/K_\tau(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \to \frac{(1/K(x))^{-\xi} - 1}{m^{-\xi} - 1} = B(x,m), \tag{9.15}$$

and, likewise, conclude for $K'_\tau(x)$ in place of $K_\tau(x)$. Therefore, $B_\tau(x,m) \to B(x,m)$ when $K(x) \neq 1$. To show that $B_\tau(x,m) \to B(x,m)$ also holds for $K(x) = 1$ with $B(x,m) = 0$, let κ' and κ'' be any positive constants such that $\kappa' < 1 < \kappa''$. By monotonicity of the quantile function, for all sufficiently small τ',

$$\frac{F_u^{-1}(\tau/\kappa'') - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \le \frac{F_u^{-1}(\tau/K_\tau(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \le \frac{F_u^{-1}(\tau/\kappa') - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}. \tag{9.16}$$

By Lemma 2(iii), as $\tau' \searrow 0$, the upper and lower bounds in (9.16) converge to

$$\frac{(1/\kappa'')^{-\xi} - 1}{m^{-\xi} - 1} \quad \text{and} \quad \frac{(1/\kappa')^{-\xi} - 1}{m^{-\xi} - 1}. \tag{9.17}$$

If in (9.17) we let $\kappa', \kappa'' \to 1$, then expressions in (9.17) $\to 0$. Therefore, since κ' and κ'' can be chosen arbitrarily close to 1, it follows from (9.16) and (9.17) that $\frac{F_u^{-1}(\tau/K_\tau(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \to 0$ as $\tau' \searrow 0$. Likewise, conclude for $K'_\tau(x)$ in place of $K_\tau(x)$. Therefore, $B_\tau(x, m) \to B(x, m) = 0$ when $K(x) = 1$.

STEP 2. By Step 1, for each $x \in \mathbf{X}$, uniformly in $(m, \tau) \in M \times \mathcal{T}(\tau')$ as $\tau' \searrow 0$,

$$B_\tau(x, m) = \frac{(x - \mu_X)'(\beta(\tau) - \beta_r)}{\mu'_X(\beta(m\tau) - \beta(\tau))} \to B(x, m). \tag{9.18}$$

Since (a) $B(x, m)$ is finite and continuous in x over $\mathbf{X}$ by conditions imposed on $K(x)$ in Condition R1, and (b) $B_\tau(x, m)$ is linear in x, the relation (9.18) also holds uniformly in $x \in \mathbf{X}$. Recall that $(x - \mu_X)_1 = 0$. Since $(x - \mu_X)_{-1}$ ranges over a nondegenerate subset of $\mathbb{R}^{d-1}$, (9.18) implies

$$\frac{\beta_{-1}(\tau) - \beta_{-1r}}{\mu'_X(\beta(m\tau) - \beta(\tau))} \to \mu(m), \tag{9.19}$$

uniformly in $(m, \tau) \in M \times \mathcal{T}(\tau')$ as $\tau' \searrow 0$, where $\mu(m)$ is some vector of finite constants. Hence, $B(x, m)$ is affine in $(x - \mu_X)$. Note also that $(x - \mu_X) = (0, x'_{-1})'$. Therefore, if $\xi = 0$, $B(x, m)$ affine and $B(x, m) = -\ln K(x)/\ln m$ imply $K(x) = e^{(x-\mu_X)'\mathbf{c}} = e^{x'_{-1}\mathbf{c}_{-1}} = e^{x'\mathbf{c}}$ for all x iff $\mathbf{c}_1 = 0$. When $\xi < 0$, $B(x, m)$ affine and $B(x, m) = (K(x)^\xi - 1)/(m^{-\xi} - 1)$ imply $K(x) = (1 + (x - \mu_X)'\mathbf{c})^{1/\xi}$, which equals $(x'\mathbf{c})^{1/\xi}$ for all x iff $\mathbf{c}_1 = 1$. Likewise, conclude for $\xi > 0$. This completes the proof of claim (i).

Claim (iii) follows directly from (9.19) and the preceding paragraph.

Claim (iv) is verified by substituting the forms of $K(x)$ found above into (9.18).

Claim (v) holds pointwise in x by Lemma 9.2(ii) and (iii). Since the left-hand side in (9.7) is linear in x and $\mathbf{X}$ is compact, it also holds uniformly in $x \in \mathbf{X}$.

A combination of Lemma 9.2(iii) with claim (v) implies claim (vi).

Claim (ii). If $\xi < 0$, by claim (iii) uniformly in k in any compact subset of $(0, \infty)$ as $T \to \infty$,

$$\begin{aligned} &a_T\left(\beta\left(\frac{k}{T}\right) - \beta_r\right) \\ &\quad \sim a_T \mathbf{c} F_u^{-1}\left(\frac{k}{T}\right) = \mathbf{c} F_u^{-1}\left(\frac{k}{T}\right)\Big/ F_u^{-1}\left(\frac{1}{T}\right) \to k^{-\xi}\mathbf{c}, \end{aligned} \tag{9.20}$$

since by Lemma 9.2(i) $F_u^{-1} \in \mathcal{R}_{-\xi}(0)$; similarly, if $\xi > 0$,

$$\begin{aligned} &a_T\left(\beta\left(\frac{k}{T}\right) - \beta_r\right) \\ &\quad \sim -a_T \mathbf{c} F_u^{-1}\left(\frac{k}{T}\right) = -\mathbf{c} F_u^{-1}\left(\frac{k}{T}\right)\Big/ F_u^{-1}\left(\frac{1}{T}\right) \to -k^{-\xi}\mathbf{c}. \end{aligned} \tag{9.21}$$

If $\xi = 0$, by $\mathbf{c}_1 = 0$, Lemma 9.2(i), (iv) and claim (iii) [using $m = e$ in $\mu(m)$], we have that uniformly in k in any compact subset of $(0, \infty)$,

$$
\begin{aligned}
&a_T\left(\beta\left(\frac{k}{T}\right) - \beta_r - b_T\mathbf{e}_1\right)\\
&\quad\sim \frac{1}{a(F^{-1}(1/T))}\\
&\qquad\times\left[\mathbf{c}\left(F_u^{-1}\left(e\frac{k}{T}\right) - F_u^{-1}\left(\frac{k}{T}\right)\right) + \mathbf{e}_1\left(F_u^{-1}\left(\frac{k}{T}\right) - F_u^{-1}\left(\frac{1}{T}\right)\right)\right]\\
&\quad\to \mathbf{c}\ln e + \mathbf{e}_1 \ln k = \mathbf{c} + \mathbf{e}_1 \ln k. \qquad \square
\end{aligned}
\tag{9.22}
$$

9.2. *Proof of Theorem* 3.1. Follows from Lemma 9.1(i). $\square$

9.3. *Proof of Theorem* 4.1.

Part 1. Referring to (2.4), notice that $Z_T(k)$ defined in (4.1) solves

$$
\widehat{Z}_T(k) \in \arg\min_{z\in\mathbb{R}^d}\left[\frac{1}{a_T}\sum_{t=1}^{T}\rho_\tau\big(a_T(U_t - b_T) - X_t'z\big)\right]
\tag{9.23}
$$

[where $z \equiv a_T(\beta - \beta_r - b_T\mathbf{e}_1)$]. Rearranging terms, the objective function becomes

$$
\begin{aligned}
&\frac{1}{a_T}\Bigg[-\tau T\overline{X}'z - \sum_{t=1}^{T}\mathbb{1}\big(a_T(U_t - b_T) \le X_t'z\big)\big(a_T(U_t - b_T) - X_t'z\big)\\
&\qquad\qquad + \tau\cdot\sum_{t=1}^{n} a_T(U_t - b_T)\Bigg].
\end{aligned}
\tag{9.24}
$$

Mutiply (9.24) by a_T and subtract

$$
\begin{aligned}
&\sum_{t=1}^{T}\mathbb{1}\big(a_T(U_t - b_T) \le -\delta\big)\big(-\delta - a_T(U_t - b_T)\big) + \sum_{t=1}^{T}\tau a_T(U_t - b_T)\\
&\qquad\qquad \text{for some } \delta > 0,
\end{aligned}
\tag{9.25}
$$

which does not affect optimization, and denote the new objective function $Q_T(z, k)$:

$$
Q_T(z, k) \equiv -\tau T\overline{X}'z + \sum_{t=1}^{T} l_\delta\big(a_T(U_t - b_T), X_t'z\big),
\tag{9.26}
$$

where

$$
l_\delta(u, v) \equiv \mathbb{1}(u \le v)(v - u) - \mathbb{1}(u \le -\delta)(-\delta - u) \qquad \text{for } \delta > 0.
\tag{9.27}
$$

Since it is a sum of convex functions in z, $Q_T(z,k)$ is convex in z. The transformations make (as shown later) Q_T a continuous functional of the point process $\widehat{\mathbf{N}}$:

$$Q_T(z,k) = -\tau T\overline{X}'z + \int_E l_\delta(j, x'z)\, d\widehat{\mathbf{N}}(j,x), \tag{9.28}$$

where the point process

$$\widehat{\mathbf{N}}(\cdot) \equiv \sum_{t\leq T} \mathbb{1}\{(a_T(U_t - b_T), X_t) \in \cdot\} \tag{9.29}$$

is taken to be a random element of the metric space $M_p(E)$ of point processes defined on the measure space $(E, \mathcal{E})$ and equipped with the metric induced by the topology of vague convergence [cf. Resnick (1987)].

It will suffice to restrict our attention to underlying measure spaces $(E, \mathcal{E})$ of the form

$$E = \begin{cases} E_1 \equiv [-\infty, \infty) \times \mathbf{X}, & \text{for type 1 tails,} \\ E_2 \equiv [-\infty, 0) \times \mathbf{X}, & \text{for type 2 tails,} \\ E_3 \equiv [0, \infty) \times \mathbf{X}, & \text{for type 3 tails,} \end{cases} \tag{9.30}$$

with σ-algebra $\mathcal{E}$ generated by the open sets of E. The topology on E_1, E_2 and E_3 is assumed to be standard so that, for example, $[-\infty, a] \times \mathbf{X}$ is compact in E_2 for $a < 0$ and in E_1 for any $a < \infty$.

Part 2 shows that, for type 1 and 3 tails, the marginal weak limit of Q_T is a finite convex function in z:

$$Q_\infty(z,k) = -k\mu_X'z + \int_E l_\delta(j, x'z)\, d\mathbf{N}(j,x), \qquad z \in \mathbb{R}^d, \tag{9.31}$$

where $\mathbf{N}$ is the Poisson point process defined in the statement of Theorem 4.1.

Part 2 also shows that, for type 2 tails, the marginal weak limit of Q_T is a finite convex function in z:

$$\begin{aligned} Q_\infty(z,k) &= -k\mu_X'z + \int_E l_\delta(j, x'z)\, d\mathbf{N}(j,x) \\ &\qquad \text{for } z \in \mathcal{Z}_N \equiv \Big\{z \in \mathbb{R}^d : \max_{x\in\mathbf{X}} x'z < 0\Big\}, \end{aligned} \tag{9.32}$$

where $\mathbf{N}$ is the Poisson point process defined in the statement of Theorem 4.1, and

$$Q_\infty(z,k) = +\infty \qquad \text{for } z \in \mathcal{Z}_P \equiv \Big\{z \in \mathbb{R}^d : \max_{x\in\mathbf{X}} x'z > 0\Big\}. \tag{9.33}$$

The function $Q_\infty(z,k)$ is convex and $l_\delta(j, x'z) = (j - x'z)^+ \geq 0$ when $j \geq -\delta$. Hence, $Q_\infty(z,k)$ is also well defined over entire $\mathcal{Z} = \{z \in \mathbb{R}^d : \max_{x\in\mathbf{X}} x'z \leq 0\}$, although it may equal $+\infty$ at $z : \max_{x\in\mathbf{X}} x'z = 0$. Also, note that $\mathcal{Z}_N \cup \mathcal{Z}_P$ is dense in $\mathbb{R}^d$.

Recall the convexity lemma [cf. Geyer (1996) and Knight (1999)], which states: Suppose (i) a sequence of convex lower-semicontinous functions $Q_T : \mathbb{R}^d \to \overline{\mathbb{R}}$ marginally converges to $Q_\infty : \mathbb{R}^d \to \overline{\mathbb{R}}$ over a dense subset of $\mathbb{R}^d$, (ii) Q_∞ is finite over a nonempty open set $\mathcal{Z}_0$, and (iii) Q_∞ is uniquely minimized at a random vector Z_∞. Then any arg min of Q_T, denoted $\widehat{Z}_T$, converges in distribution to Z_∞.

We showed (i) and (ii) in Step 2, and we assumed (iii). (A sufficient condition for uniqueness is given in Remark 4.4.) Hence, application of the convexity lemma to our case gives

$$\widehat{Z}_T(k) \xrightarrow{d} Z_\infty(k) \equiv \arg\min_{z \in \mathbb{R}^d} Q_\infty(z, k). \tag{9.34}$$

Note also that, for type 2, tails, the arg min $Z_\infty(k)$ necessarily belongs to $\mathcal{Z} = \{z \in \mathbb{R}^d : \max_{x \in \mathbf{X}} x'z \le 0\}$. This gives us the conclusion stated in Theorem 4.1 upon noting that $Q_\infty(z, k)$ differs from the limit objective function of Theorem 4.1 only by a finite random variable that does not depend on z.

Part 2. It remains to verify that (I) there exists a nonempty open set $\mathcal{Z}_0$ such that $Q_\infty(z, k)$ is finite a.s. for all $z \in \mathcal{Z}_0$ and (II) $Q_\infty(\cdot, k)$ is, indeed, the weak marginal limit of $Q_T(\cdot, k)$.

To show (I), when tails are of type 1 and 3, choose $\mathcal{Z}_0$ as any open bounded subset of $\mathbb{R}^d$; when tails are of type 2, additionally require $\mathcal{Z}_0 \subset \mathcal{Z}_N$ for each l (possible by compactness of $\mathbf{X}$). For any $z \in \mathcal{Z}_0$, $(u, x) \mapsto l_\delta(u, x'z)$ is in $C_K(E)$ (continuous functions on E vanishing outside a compact set K) by the arguments in (II). This implies $\int_E l_\delta(u, x'z)\, d\mathbf{N}(u, x)$ is finite a.s., since $\mathbf{N} \in M_p(E)$.

To show (II), $Q_\infty(\cdot, k)$ is the marginal weak limit of $\{Q_T(\cdot, k)\}$ iff for any finite collection $(z_j, j = 1, \ldots, l)$, $(Q_T(z_j, k), j = 1, \ldots, l) \xrightarrow{d} (Q_\infty(z_j, k), j = 1, \ldots, l)$. Since $\overline{X}'z_j \xrightarrow{p} \mu_X' z_j$ and $\tau T \to k > 0$, it remains to verify

$$\begin{aligned} &\left(\int_E l_\delta(u, x'z_j)\, d\widehat{\mathbf{N}}(u, x),\ j = 1, \ldots, l\right) \\ &\qquad \xrightarrow{d} \left(\int_E l_\delta(u, x'z_j)\, d\mathbf{N}(u, x),\ j = 1, \ldots, l\right). \end{aligned} \tag{9.35}$$

Define the mapping $\mathbf{T} : M_p(E) \to \mathbb{R}^l$ (for $E = E_1$, E_2 or E_3) by

$$\mathbf{T} : N \mapsto \left(\int_E l_\delta(u, x'z_j)\, dN(u, x),\ j = 1, \ldots, l\right). \tag{9.36}$$

(a) Consider type 1 tails and set $E = E_1$. The map $(u, x) \mapsto l_\delta(u, x'z_j)$ is in $C_K(E_1)$ (continuous functions on E_1 vanishing outside a compact set K), since by construction it is continuous on E_1 and vanishes outside $K \equiv [-\infty, \max(\kappa, -\delta)] \times \mathbf{X}$, where $\kappa = \max_{x \in \mathbf{X}, z \in \{z_1, \ldots, z_l\}} x'z$. K is compact in E_1 since $\kappa < \infty$ by Condition R2. Hence, $\widehat{\mathbf{N}} \mapsto \mathbf{T}(\widehat{\mathbf{N}})$ is continuous from $M_p(E_1)$ to $\mathbb{R}^l$. Thus, $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E_1)$ implies $\mathbf{T}(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}(\mathbf{N})$.

(b) Consider type 3 tails and set $E = E_3$. The map $(u, x) \mapsto l(u, x'z_j)$ is in $C_K(E_3)$: by construction, it is continuous on E_3 and vanishes outside $K \equiv [0, \max(\kappa, 0)] \times \mathbf{X}$, where $\kappa = \max_{x \in \mathbf{X}, z \in \{z_1, \dots, z_l\}} x'z$. K is compact in E_3 since $\kappa < \infty$ by Condition R2. Therefore, $\widehat{\mathbf{N}} \mapsto \mathbf{T}(\widehat{\mathbf{N}})$ is continuous from $M_p(E_3)$ to $\mathbb{R}^l$. Hence, $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E_3)$ implies $\mathbf{T}(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}(\mathbf{N})$.

(c) Consider type 2 tails and set $E = E_2$. (c)(i) shows that (9.35) holds on $\mathcal{Z}_N$, while (c)(ii) shows that $Q_n(z) \xrightarrow{p} \infty$ for any $z \in \mathcal{Z}_P$. [Sets $\mathcal{Z}_N$ and $\mathcal{Z}_P$ are defined in (9.32) and (9.33).]

(i) The map $(u, x) \mapsto l_\delta(u, x'z)$ is in $C_K(E_2)$ if $z \in \mathcal{Z}_N$, since, by construction, it is continuous on E_2 and vanishes outside $K \equiv [-\infty, \max(\kappa, -\delta)] \times \mathbf{X}$, where $\kappa = \max_{x \in \mathbf{X}, z \in \{z_1, \dots, z_l\}} x'z$. K is compact in E_2 since $\kappa < 0$ if $z \in \mathcal{Z}_N$. Hence, $\widehat{\mathbf{N}} \mapsto \mathbf{T}(\widehat{\mathbf{N}})$ is continuous from $M_p(E_2)$ to $\mathbb{R}^l$. Then $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E_2)$ implies $\mathbf{T}(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}(\mathbf{N})$.

(ii) Observe that $I \equiv \sum_{t \le T} l_\delta(a_T U_t, X_t'z) \mathbb{1}(a_T U_t \le -\delta) = O_p(1)$ by the argument in (i). Observe that $l_\delta(u, v) = (v - u)^+ \ge 0$ for any $u \ge -\delta$. Hence,

$$
l_\delta(u, v) = \mathbb{1}(-\delta \le u \le v)(v - u) \ge \mathbb{1}(-\delta \le u \le 0, v \ge \varepsilon)\varepsilon \quad \text{for any } u \ge -\delta \text{ and any } \varepsilon > 0. \tag{9.37}
$$

For a given $z \in \mathcal{Z}_P$, since $\mathbf{X}$ equals the support of X, $\max_{x \in \mathbf{X}} x'z > 0$ implies that $X'z \ge \varepsilon$ occurs with positive probability for some $\varepsilon > 0$. Fix this ε. Since $1/a_T \to \infty$ for type 2 tails, $P(-\delta/a_T \le U \le 0, X'z \ge \varepsilon) \to \pi = P(U \le 0, X'z \ge \varepsilon) > 0$. $\pi > 0$ because $\inf_{x \in \mathbf{X}} P(U \le 0 | X = x) > 0$ for type 2 tails by assumptions in Condition R1. Therefore, $II \equiv \sum_{t \le T} \mathbb{1}(-\delta/a_T \le U_i \le 0, X_i'z \ge \varepsilon)\varepsilon \xrightarrow{p} +\infty$ in $\overline{\mathbb{R}}$. Since $Q_T(z, k) \ge -k\mu_X'z + I + II$ by (9.37), $Q_T(z, k) \xrightarrow{p} +\infty$ for any $z \in \mathcal{Z}_P$.

Part 3. By Lemma 9.1(ii), $a_T(\beta(\tau) - \beta_r - b_T \mathbf{e}_1) \to \eta(k)$. Hence, $\widehat{Z}_T^c(k) \xrightarrow{d} Z_\infty^c(k) \equiv Z_\infty(k) - \eta(k)$.

Part 4. $(\widehat{Z}_T(k_j)', j = 1, \dots, l)' \in \arg\min_{z \in \mathbb{R}^{d \times l}} [Q_T(z_1, k_1) + \cdots + Q_T(z_l, k_l)]$, for $z = (z_1, \dots, z_l)$. Since this objective is a sum of objective functions in Parts 1 and 2, the previous derivation of the marginal limit and subsequent arguments apply very similarly to $Q_T(z_1, k_1) + \cdots + Q_T(z_l, k_l)$ to conclude that $(\widehat{Z}_T(k_j)', j = 1, \dots, l)' \xrightarrow{d} (\widehat{Z}_\infty(k_j)', j = 1, \dots, l)' = \arg\min_{z \in \mathbb{R}^{d \times l}} [Q_\infty(z_1, k_1) + \cdots + Q_\infty(z_l, k_l)]$. $\square$

9.4. *Weak limit of* $\widehat{\mathbf{N}}$.

LEMMA 9.3 [Resnick (1987), Proposition 3.22]. *Suppose* $\mathbf{N}$ *is a simple point process in* $M_p(E)$, $\mathcal{T}$ *is a basis of relatively compact open sets such* *that* $\mathcal{T}$ *is closed under finite unions and intersections and, for any* $F \in \mathcal{T}$, $P(\mathbf{N}(\partial F) = 0) = 1$. *Then* $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ *in* $M_p(E)$ *if, for all* $F \in \mathcal{T}$,

$$\lim_{T\to\infty} P[\widehat{\mathbf{N}}(F) = 0] = P[\mathbf{N}(F) = 0], \tag{9.38}$$

$$\lim_{T\to\infty} E\widehat{\mathbf{N}}(F) = E\mathbf{N}(F) < \infty. \tag{9.39}$$

REMARK 9.1. In our case, $\mathcal{T}$ consists of finite unions and intersections of bounded open rectangles in E_1, E_2 and E_3 [cf. Resnick (1987)].

We impose Meyer (1973) conditions on the "rare" events $A_t^T(F) \equiv \{w \in \Omega : (a_T(U_t - b_T), X_t) \in F\}$.

LEMMA 9.4 (Poisson limits under Meyer mixing conditions). *Suppose that, for any* $F \in \mathcal{T}$, *the triangular sequence of events* $\{(A_t^T(F), t \le T), T \ge 1\}$ *is stationary and* α*-mixing with mixing coefficient* $\alpha_T(\cdot)$, *condition* (9.39) *holds, and the Meyer type condition holds*: *There exist sequences of integers* $(p_n, n \ge 1)$, $(q_n, n \ge 1)$, $(t_n = n(p_n + q_n), n \ge 1)$ *such that as* $n \to \infty$, *for some* $r > 0$, (a) $n^r \alpha_{t_n}(q_n) \to 0$, (b) $q_n/p_n \to 0$, $p_{n+1}/p_n \to 1$, *and* (c) $I_{p_n} = \sum_{i=1}^{p_n-1}(p_n - i)P(A_1^{t_n}(F) \cap A_{i+1}^{t_n}(F)) = o(1/n)$. *Then in* $M_p(E)$, $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$, *a Poisson point process with mean measure* $m : m(F) \equiv \lim_{T\to\infty} E\widehat{\mathbf{N}}(F)$.

PROOF. For any $F : m(F) > 0$, $\lim_{T\to\infty} P[\widehat{\mathbf{N}}(F) = 0] = P[\mathbf{N}(F) = 0] = e^{-m(F)}$, by Meyer (1973). The same also holds for $F : m(F) = 0$, since $E\widehat{\mathbf{N}}(F) \to 0$ implies $P(\widehat{\mathbf{N}}(F) = 0) \to 1$. Conclude by Lemma 9.3. $\square$

REMARK 9.2. Condition $I_{P_n} = o(1/n)$ prevents clusters of "rare" events $A_t^T(F)$, eliminating compound Poisson processes as limits.

LEMMA 9.5 (Limit $\mathbf{N}$ under Conditions R1 and R2). *Suppose Conditions* R1 *and* R2 *hold and that* (Y_t, X_t) *is an i.i.d. or stationary strongly mixing sequence that satisfies the conditions of Lemma* 9.4 *with* (a_T, b_T) *defined in* (4.2). *Then*:

(i) $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ *in* $M_p(E)$, *where* $E = E_1, E_2$ *and* E_3 *for tails of types* 1, 2 *and* 3, *respectively.* $\mathbf{N}$ *is a Poisson point process with mean intensity measure*: $m(du, dx) = K(x) \times dh(u) \times dF_X(x)$, *where* $h(u) = e^u$ *for type* 1, $h(u) = (-u)^{-1/\xi}$ *for type* 2, *and* $h(u) = u^{-1/\xi}$ *for type* 3 *tails.*

(ii) *Points* $(J_i, \mathcal{X}_i)$ *of* $\mathbf{N}$ *have the representation* $(J_i, \mathcal{X}_i, i \ge 1) \stackrel{d}{=} (h^{-1}(\Gamma_i / K(\mathcal{X}_i)), \mathcal{X}_i, i \ge 1)$, *where* h^{-1} *is the inverse of* h, $\Gamma_i = \mathcal{E}_1 + \cdots + \mathcal{E}_i, i \ge 1$ ($\{\mathcal{E}_i\}$ *are i.i.d. standard exponential*), *and* $\{\mathcal{X}_i\}$ *are i.i.d. r.v.s with law* F_X, *independent of* $\{\mathcal{E}_i\}$.

PROOF. To show (i), by Lemmas 9.3 and 9.4 the proof reduces to verifying $\lim_T E\widehat{\mathbf{N}}(F) = m(F)$ for all F in $\mathcal{T}$. For example, as in Leadbetter, Lindgren and Rootzén [(1983), page 103], it suffices to consider F of the form $F = \bigcup_{j=1}^k F_j$, where $F_j = (l_j, u_j) \times \mathbf{X}_j$, where $F_1, \dots, F_k$ are nonoverlapping, nonempty subsets of E, and $\mathbf{X}_1, \dots, \mathbf{X}_k$ are intersections of open bounded rectangles of $\mathbb{R}^d$ with $\mathbf{X}$. Then by the stationarity and F_j's nonoverlapping,

$$
\begin{aligned}
E\widehat{\mathbf{N}}(F) &= E\sum_{t=1}^{T} \mathbb{1}\big[\big(a_T(U_t - b_T), X_t\big) \in F\big] \\
&= \sum_{j=1}^{k} T P\big[\big(a_T(U - b_T), X\big) \in (l_j, u_j) \times \mathbf{X}_j\big] \\
&= \sum_{j=1}^{k} T \cdot E\big(P\big[\big(a_T(U - b_T), X\big) \in (l_j, u_j) \times \mathbf{X}_j | X\big]\big) \\
&= \sum_{j=1}^{k} T \cdot E\big(P\big[\big(a_T(U - b_T) \in (l_j, u_j)\big| X\big] \cdot \mathbb{1}[X \in \mathbf{X}_j]\big) \\
&= \sum_{j=1}^{k} T \cdot E\big((F_U[u_j/a_T + b_T | X] \\
&\qquad\qquad - F_U[l_j/a_T + b_T | X]) \cdot \mathbb{1}[X \in \mathbf{X}_j]\big).
\end{aligned} \tag{9.40}
$$

Suppose that $l_j > -\infty$ for all j. Then as $T \to \infty$,

$$
\begin{aligned}
E\widehat{\mathbf{N}}(F) &= \sum_{j=1}^{k} E\bigg(\bigg(\frac{F_U[u_j/a_T + b_T | X]}{F_u[u_j/a_T + b_T]} \cdot T \cdot F_u[u_j/a_T + b_T] \\
&\qquad - \frac{F_U[l_j/a_T + b_T | X]}{F_u[l_j/a_T + b_T]} \cdot T \cdot F_u[l_j/a_T + b_T]\bigg) \cdot \mathbb{1}[X \in \mathbf{X}_j]\bigg) \\
&\sim \sum_{j=1}^{k} E\big(\big(K(X)[h(u_j) - h(l_j)]\big)\mathbb{1}[X \in \mathbf{X}_j]\big) \qquad (9.41) \\
&= \sum_{j=1}^{k} \int_{F_j} K(x)\, dh(u) \times dF_X(x) \\
&= \sum_{j=1}^{k} m(F_j) = m(F).
\end{aligned}
$$

In (9.41), $\sim$ follows from two observations. First, the assumed tail equivalence Condition R1 implies

$$\frac{F_U[l/a_T + b_T|x]}{F_u[l/a_T + b_T]} \sim K(x) \qquad \text{uniformly in } x \in \mathbf{X}, \tag{9.42}$$

since by definition of (a_T, b_T) given in (4.2), $l/a_T + b_T \searrow F_u^{-1}(0) = 0$ or $= -\infty$ for any $l \in (-\infty, \infty)$ for type 1 tails, any $l \in (-\infty, 0)$ for type 2 tails, and $l \in [0, \infty)$ for type 3 tails. Second, for example, as in Leadbetter, Lindgren and Rootzén [(1983), page 103], the definition of the tail types (3.1) implies that (a) for tails of type 2, for any $l < 0$, $TF_u(l/a_T) = TF_u(-lF_u^{-1}(\frac{1}{T})) \sim (-l)^{-1/\xi} TF_u(F_u^{-1}(\frac{1}{T})) \sim (-l)^{-1/\xi}$, (b) for tails of type 3, for any $l > 0$, $TF_u(l/a_T) = TF_u(lF_u^{-1}(\frac{1}{T})) \sim l^{-1/\xi} TF_u(F_u^{-1}(\frac{1}{T})) \sim l^{-1/\xi}$ and (c) for tails of type 1, for any $l \in \mathbb{R}$, $TF_u(l/a_T + b_T) = TF_u(l/a(F_u^{-1}(\frac{1}{T})) + F_u^{-1}(\frac{1}{T})) \sim e^l TF_u(F_u^{-1}(\frac{1}{T})) \sim e^l$.

On the other hand, if for some j's, $l_j = -\infty$ for type 1 or 2 tails, then we have the replacement $TF_U[l_j/a_T + b_T|X] = 0$ in (9.40), and (9.41) follows similarly.

To show (ii), construct a Poisson random measure (PRM) with the given $m(\cdot)$. First, define a canonical homogeneous PRM $\mathbf{N}_1$ with points $\{\Gamma_i, i \geq 1\}$. It has the mean measure $m_1(du) = du$ on $[0, \infty)$, for example, Resnick (1987). Second, by Proposition 3.8 in Resnick (1987), the composed point process $\mathbf{N}_2$ with points $\{\Gamma_i, \mathcal{X}_i\}$ is PRM with mean measure $m_2(du, dx) = du \times dF_X(x)$ on $[0, \infty) \times \mathbf{X}$, because $\{\mathcal{X}_i\}$ are i.i.d. and are independent of $\{\Gamma_i\}$. Finally, the point process $\mathbf{N}$ with the transformed points $\{\mathbf{T}(\Gamma_i, \mathcal{X}_i)\}$, where $\mathbf{T}: (u, x) \mapsto (h^{-1}(u/K(x)), x)$, is PRM with the desired mean measure on $E \times \mathbf{X}$, $m(dj, dx) = m_2 \circ \mathbf{T}^{-1}(dj, dx) = K(x) \times dh(j) \times dF_X(x)$, by Proposition 3.7 in Resnick (1987). $\square$

9.5. *Proof of Theorem* 5.1. Step 1 outlines the overall proof using standard convexity arguments, while the *main* Step 2 invokes regular variation assumptions on the conditional quantile density to demonstrate a quadratic approximation of the criterion function. Step 3 shows joint convergence of several regression quantile statistics. Step 4 demonstrates that a_T can be estimated consistently.

STEP 1. With reference to (2.4), notice that $\widehat{Z}_T \equiv a_T(\hat{\beta}(\tau) - \beta(\tau))$, defined in (5.2), minimizes

$$Q_T(z, \tau) \equiv \frac{a_T}{\sqrt{\tau T}} \sum_{t=1}^{T} \biggl(\rho_\tau \biggl(Y_t - X_t'\beta(\tau) - \frac{X_t' z}{a_T} \biggr) - \rho_\tau \bigl(Y_t - X_t'\beta(\tau) \bigr) \biggr). \tag{9.43}$$

Using Knight's identity,

$$\begin{aligned} &\rho_\tau(u - v) - \rho_\tau(u) \\ &\quad = -v\bigl(\tau - \mathbb{1}(u < 0)\bigr) + \int_0^v \bigl(\mathbb{1}(u \leq s) - \mathbb{1}(u \leq 0)\bigr)\, ds, \end{aligned} \tag{9.44}$$

write, a.s.,

$$
\begin{aligned}
&Q_T(z,\tau) = W_T(\tau)'z + G_T(z,\tau),\\
&W_T(\tau) \equiv \frac{-1}{\sqrt{\tau T}} \sum_{t=1}^{T} \big(\tau - \mathbb{1}[Y_t < X_t'\beta(\tau)]\big) X_t,\\
&G_T(z,\tau) \equiv \frac{a_T}{\sqrt{\tau T}} \Bigg(\sum_{t=1}^{T} \int_0^{X_t'z/a_T} \big[\mathbb{1}\big(Y_t - X_t'\beta(\tau) \le s\big)\\
&\qquad\qquad - \mathbb{1}\big(Y_t - X_t'\beta(\tau) \le 0\big)\big]\, ds \Bigg).
\end{aligned}
\tag{9.45}
$$

By Lemma 9.6, $W_T(\tau) \xrightarrow{d} W \equiv N(0, EXX')$, and by Step 2,

$$
G_T(z,\tau) \xrightarrow{p} \frac{1}{2}\left(\frac{m^{-\xi}-1}{-\xi}\right) z'\mathcal{Q}_H z, \qquad m > 1,
\tag{9.46}
$$

where $\mathcal{Q}_H \equiv E[H(X)]^{-1}XX'$, $H(x) \equiv x'\mathbf{c}$ for type 2 and 3 tails, and $H(x) = 1$ for type 1 tails. Thus, the weak marginal limit of $Q_T(z)$ is given by

$$
Q_\infty(z) = W'z + \frac{1}{2} \cdot \left(\frac{m^{-\xi}-1}{-\xi}\right) \cdot z'\mathcal{Q}_H z.
\tag{9.47}
$$

We have that EXX' is positive definite and by Theorem 3.1 that $0 < H(X) < c < \infty$ for some constant c. Thus, $\mathcal{Q}_H$ is finite and $\mathcal{Q}_H$ is positive definite. Indeed, $z'\mathcal{Q}_H z = E(X'z)^2/H(X) = 0$ for some $z \ne 0$ if and only if $X'z = 0$ a.s., which contradicts EXX' positive definite. Thus, the marginal limit $Q_\infty(z)$ is uniquely minimized at $Z_\infty \equiv (\frac{\xi}{m^{-\xi}-1})\mathcal{Q}_H^{-1}W = N(0, \frac{\xi^2}{(m^{-\xi}-1)^2}\mathcal{Q}_H^{-1}EXX'\mathcal{Q}_H^{-1})$. By the convexity lemma [e.g., Geyer (1996) and Knight (1999)], $\widehat{Z}_T \xrightarrow{d} Z_\infty$.

STEP 2. This step demonstrates that as $\tau \searrow 0$,

$$
EG_T(z,\tau) \to \frac{1}{2} \cdot \left(\frac{m^{-\xi}-1}{-\xi}\right) \cdot z'\mathcal{Q}_H z,
\tag{9.48}
$$

while Lemma 9.6 shows that $\operatorname{Var}(G_T(z,\tau)) \to 0$. In what follows, F_t, f_t and E_t denote $F_U(\cdot|X_t)$, $f_U(\cdot|X_t)$ and $E[\cdot|X_t]$, respectively, where U is the auxiliary error constructed in Condition R1.

Since

$$
\begin{aligned}
&G_T(z,\tau)\\
&\quad\equiv \sum_{t=1}^{T} a_T \cdot \left(\int_0^{X_t'z/a_T} \left[\frac{\mathbb{1}(Y_t - X_t'\beta(\tau) \le s) - \mathbb{1}(Y_t - X_t'\beta(\tau) \le 0)}{\sqrt{\tau T}} \right] ds \right)\\
&\quad= \sum_{t=1}^{T} \left(\int_0^{X_t'z} \left[\frac{\mathbb{1}(Y_t - X_t'\beta(\tau) \le s/a_T) - \mathbb{1}(Y_t - X_t'\beta(\tau) \le 0)}{\sqrt{\tau T}} \right] ds \right),
\end{aligned}
\tag{9.49}
$$

we have

$$
\begin{aligned}
EG_T(z,\tau) &= T \cdot E\left(\int_0^{X_t'z} \frac{F_t[F_t^{-1}(\tau)+s/a_T]-F_t[F_t^{-1}(\tau)]}{\sqrt{\tau T}}\,ds\right)\\
&\overset{(1)}{=} T \cdot E\left(\int_0^{X_t'z} \frac{f_t\{F_t^{-1}(\tau)+o(F_u^{-1}(m\tau)-F_u^{-1}(\tau))\}}{a_T \cdot \sqrt{\tau T}} \cdot s \cdot ds\right)\\
&\overset{(2)}{\sim} T \cdot E\left(\int_0^{X_t'z} \frac{f_t\{F_t^{-1}(\tau)\}}{a_T \cdot \sqrt{\tau T}} \cdot s \cdot ds\right)\\
&= T \cdot E\left(\frac{1}{2} \cdot (X_t'z)^2 \cdot \frac{f_t\{F_t^{-1}(\tau)\}}{a_T \cdot \sqrt{\tau T}}\right) \qquad (9.50)\\
&= E\left(\frac{1}{2} \cdot (X_t'z)^2 \cdot \frac{F_u^{-1}(m\tau)-F_u^{-1}(\tau)}{\tau(f_t\{F_t^{-1}(\tau)\})^{-1}}\right)\\
&\overset{(3)}{\sim} E\left(\frac{1}{2} \cdot (X_t'z)^2 \cdot \frac{1}{H(X)} \cdot \frac{m^{-\xi}-1}{-\xi}\right)\\
&\equiv \frac{1}{2} \cdot \frac{m^{-\xi}-1}{-\xi} \cdot z'\mathcal{Q}_H z.
\end{aligned}
$$

Equality (1) is by the definition of a_T and a Taylor expansion. Indeed, since $\tau T \to \infty$ uniformly over s in any compact subset of $\mathbb{R}$,

$$s/a_T = s \cdot \big(F_u^{-1}(m\tau)-F_u^{-1}(\tau)\big)/\sqrt{\tau T} = o\big(F_u^{-1}(m\tau)-F_u^{-1}(\tau)\big). \tag{9.51}$$

To show equivalence (2), it suffices to prove that, for any sequence $v_\tau = o(F_u^{-1}(m\tau)-F_u^{-1}(\tau))$ with $m>1$ as $\tau \searrow 0$,

$$f_t\big(F_t^{-1}(\tau)+v_\tau\big) \sim f_t\big(F_t^{-1}(\tau)\big) \qquad \text{uniformly in } t. \tag{9.52}$$

This will be shown by using the assumption made in Condition R3, which is that uniformly in t, $1/f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(X_t))/\partial\tau$, where $\partial F_u^{-1}(\tau)/\partial\tau$ is regularly varying with index $-\xi-1$.

To be clear, let us first show (9.52) for the special case of $f_t = f_u$ and $F_t^{-1}(\tau) = F_u^{-1}(\tau)$:

$$f_u\big(F_u^{-1}(\tau)+v_\tau\big) \sim f_u\big(F_u^{-1}(\tau)\big). \tag{9.53}$$

By the regular variation property of $\partial F_u^{-1}(\tau)/\partial\tau = 1/f_u(F_u^{-1}(\tau))$, locally uniformly in l [uniformly in l in any compact subset of $(0,\infty)$],

$$f_u\big(F_u^{-1}(l\tau)\big) \sim l^{\xi+1} f_u\big(F_u^{-1}(\tau)\big). \tag{9.54}$$

That is, locally uniformly in l,

$$f_u\big(F_u^{-1}(\tau)+[F_u^{-1}(l\tau)-F_u^{-1}(\tau)]\big) \sim l^{\xi+1} f_u\big(F_u^{-1}(\tau)\big). \tag{9.55}$$

Hence, for any $l_\tau \to 1$,

$$f_u\big(F_u^{-1}(\tau) + [F_u^{-1}(l_\tau \tau) - F_u^{-1}(\tau)]\big) \sim f_u\big(F_u^{-1}(\tau)\big). \tag{9.56}$$

Hence, for any sequence $v_\tau = o([F^{-1}(m\tau) - F^{-1}(\tau)])$ with $m > 1$ as $\tau \searrow 0$,

$$f_u\big(F^{-1}(\tau) + v_\tau\big) \sim f_u\big(F_u^{-1}(\tau)\big), \tag{9.57}$$

because for any such $\{v_\tau\}$, in view of Lemma 9.2(iii), we can choose a sequence $\{l_\tau\}$ such that $\{v_\tau\} = \{[F_u^{-1}(l_\tau \tau) - F_u^{-1}(\tau)]\}$ and $l_\tau \to 1$ as $\tau \searrow 0$.

Next, let us strengthen the claim (9.53) to (9.52), completing the proof of equivalence (2) in (9.50). Since

(a) $1/f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(X_t))/\partial\tau = 1/\{K(X_t) f_u[F_u^{-1}(\tau/K(X_t))]\}$ uniformly in t by Condition R3, and

(b) $f_u(F_u^{-1}(l\tau/K)) \sim (l/K)^{\xi+1} f_u(F_u^{-1}(\tau)) \sim (l)^{\xi+1} f_u(F_u^{-1}(\tau/K))$, locally uniformly in l and uniformly in $K \in \{K(x) : x \in \mathbf{X}\}$ [compact by assumptions on $K(\cdot)$ and $\mathbf{X}$], by (9.54) we have that locally uniformly in l and uniformly in t,

$$f_t\big(F_t^{-1}(l\tau)\big) \sim l^{\xi+1} f_t\big(F_t^{-1}(\tau)\big). \tag{9.58}$$

Repeating the steps (9.55)–(9.57) with $f_t(F_t^{-1}(l\tau))$ in place of $f_u(F_u^{-1}(l\tau))$, we obtain the required conclusion (9.52).

The equivalence (3) in (9.50) can be shown as follows. By (a), uniformly in t,

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_t[F_t^{-1}(\tau)])^{-1}} \sim \frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(K(X_t) f_u[F_u^{-1}(\tau/K(X_t))])^{-1}}. \tag{9.59}$$

By (b) we have that uniformly in t,

$$f_u\big[F_u^{-1}\big(\tau/K(X_t)\big)\big] \sim \big(1/K(X_t)\big)^{\xi+1} \cdot f_u\big(F^{-1}(\tau)\big). \tag{9.60}$$

Putting (9.59) and (9.60) together, we have uniformly in t,

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_t[F_t^{-1}(\tau)])^{-1}} \sim \frac{1}{K(X_t)^\xi} \cdot \frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_u[F_u^{-1}(\tau)])^{-1}} \tag{9.61}$$

$$= \frac{1}{H(X_t)} \cdot \frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_u[F_u^{-1}(\tau)])^{-1}}, \tag{9.62}$$

where $H(X_t) = X_t'\mathbf{c}$ for $\xi \neq 0$ and $H(X_t) = 1$ for $\xi = 0$. Finally, by the regular variation property, (9.54),

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_u[F_u^{-1}(\tau)])^{-1}} \equiv \int_1^m \frac{f_u[F_u^{-1}(\tau)]}{f_u[F_u^{-1}(s\tau)]}\, ds \tag{9.63}$$

$$\sim \int_1^m s^{-\xi-1}\, ds \tag{9.64}$$

$$= \frac{m^{-\xi} - 1}{-\xi} \qquad (\ln m \text{ if } \xi = 0). \tag{9.65}$$

Putting (9.61)–(9.65) together gives (3) in (9.50).

STEP 3. For $\widehat{Z}_T(l)$ defined in (5.3), notice that $(\widehat{Z}_T(l_i), i = 1, \ldots, k) \in \arg\min_{z \in \mathbb{R}^{d\times k}}[Q_T(z_1, l_1\tau) + \cdots + Q_T(z_k, l_k\tau)] = \arg\min_{z\in\mathbb{R}^{d\times k}}[\sum_{i=1}^k W_T(\tau l_i)' \times z_i + G_T(z_i, \tau l_i)]$ for $z = (z_1', \ldots, z_k')'$, where the functions $Q_T(\cdot,\cdot)$, $W_T(\cdot)$ and $G_T(\cdot,\cdot)$ are defined in (9.45). Since this objective function is a sum of the objective functions in the preceding steps, it retains the properties of the elements summed. Therefore, the previous argument applies to conclude that the marginal limit of this objective function is given by $\sum_{i=1}^k W(l_i)'z_i + G(z_i, l_i)$, where $(W(l_i), i \le k) \equiv N(0, \Sigma)$ with $EW(l_i)W(l_j)' \equiv EXX' \min(l_i, l_j)/\sqrt{l_i l_j}$ and, by calculations that are identical to those in the preceding section, $G(z, l_i) \equiv G(z) \equiv \frac{1}{2}\cdot(\frac{m^{-\xi}-1}{-\xi})\cdot z'\mathcal{Q}_H z$. The limit objective function is minimized at $(Z_\infty(l_i), i \le k) = (\frac{\xi}{m^{-\xi}-1}\cdot \mathcal{Q}_H^{-1} W(l_i), i \le k)$. Therefore, $(\widehat{Z}_T(l_i), i \le k) \xrightarrow{d} (Z_\infty(l_i), i \le k)$.

STEP 4. It suffices to prove the result for $l = 1$. Then

$$\begin{aligned} &\frac{\overline{X}'(\hat\beta(m\tau) - \hat\beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \\ &\qquad \equiv \frac{\overline{X}'(\hat\beta(m\tau) - \beta(m\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \\ &\qquad\quad - \frac{\overline{X}'(\hat\beta(\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} + \frac{\overline{X}'(\beta(m\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \xrightarrow{p} 1, \end{aligned} \tag{9.66}$$

since the first two elements on the right-hand side are $O_p(\frac{1}{\sqrt{\tau T}}) = o_p(1)$ by the first part of Theorem 5.1. □

9.6. *CLT for $W_T(\tau)$ and LLN for $G_T(z, \tau)$.*

LEMMA 9.6 (CLT and LLN). *Let $\{Y_j, X_j\}_{-\infty}^t$ be an i.i.d. or a stationary α-mixing sequence. The following statements are true for $W_T(\cdot)$ and $G_T(\cdot,\cdot)$, defined in* (9.45), *as $\tau \searrow 0$ and $\tau T \to \infty$*:

(i) *Suppose mixing coefficients satisfy $\alpha_j = O(j^{-\phi})$ with $\phi > 2$, and for any K sufficiently close to 0^+ or $-\infty$, uniformly in t and $s \ge 1$, and some $C > 0$ [P_t denotes $P(\cdot|\mathcal{F}_t)$, $\mathcal{F}_t \equiv \sigma(\{Y_j, X_j\}_{-\infty}^{t-1})$]*

$$P_t(U_t \le K, U_{t+s} \le K) \le C P_t(U_t \le K)^2. \tag{9.67}$$

Then for any finite collection of positive constants $l_1, \ldots, l_m$,

$$\{W_T(\tau l_1)', \ldots, W_T(\tau l_m)'\}' \xrightarrow{d} \big(W(l_1)', \ldots, W(l_k)'\big)' = N(0, \Sigma)$$
$$\text{with } EW(l_i)W(l_j)' \equiv EXX' \min(l_i, l_j)/\sqrt{l_i l_j}.$$

(ii) *If, in addition*, $\alpha_j = O(j^{-\phi})$ *with* $\phi > \frac{1}{1-\gamma}$ *for* $0 < \gamma < 1$ *and* $\tau^{1-2/\gamma}/T \to 0$, *then*

$$\operatorname{Var}\bigl(G_T(z,\tau)\bigr) \to 0. \tag{9.68}$$

REMARK 9.3. In the i.i.d. case the claim (i) simply follows from the Lindeberg–Feller CLT. In the dependent case condition (9.67) requires that the extremal events should not cluster, which leads to the same limits as under i.i.d. sampling. This condition may possibly be refined along the lines of Watts, Rootzén and Leadbetter (1982), who dealt with the nonregression case. (9.67) is analogous to the no-clustering conditions of Robinson [(1983), A7.4, page 191] used in the context of kernel estimation.

PROOF OF LEMMA 9.6. To show (i), $\{W_T(\tau l_i)', i \le m\}'$ suits the CLT of Robinson (1983), which implies the same weak limit as under i.i.d. sampling. His conditions A7.1 (with $q = 0$), A7.2 and A7.3 are satisfied automatically. The assumed above mixing condition implies $\sum_{j=1}^{\infty} j\alpha_j < \infty$, which implies his condition A3.3. Last, condition (9.67) immediately implies his condition A7.4.

To show (ii), suppress τ. Then from (9.49),

$$\operatorname{Var}\bigl(G_T(z)\bigr) = \tau^{-1}\left(\operatorname{Var}(\lambda_1) + 2\sum_{k=1}^{T-1}\frac{T-k}{T}\operatorname{Cov}(\lambda_1,\lambda_{1+k})\right),$$

for

$$\lambda_t = \int_0^{X_t'z}\bigl[\mathbb{1}\bigl(Y_t - X_t'\beta(\tau) \le s/a_T\bigr) - \mathbb{1}\bigl(Y_t - X_t'\beta(\tau) \le 0\bigr)\bigr]\,ds.$$

By Condition R2, $|\lambda_t| \le K_0|\mu_t|$, for

$$\mu_t = \bigl(\mathbb{1}\bigl(Y_t - X_t'\beta(\tau) \le X_t'z/a_T\bigr) - \mathbb{1}\bigl(Y_t - X_t'\beta(\tau) \le 0\bigr)\bigr)$$

and some $K_0 < \infty$. Hence,

$$\begin{aligned}\operatorname{Var}(\lambda_1) &= O(E\lambda_1^2) \overset{(1)}{=} O(E\mu_1^2) \overset{(2)}{=} O(E|\mu_1|)\\ &\overset{(3)}{=} O\bigl(f_u(F_u^{-1}(\tau))a_T^{-1}\bigr) = O\bigl(\sqrt{\tau/T}\bigr),\end{aligned} \tag{9.69}$$

where (1) is by $|\lambda_t| \le K_0|\mu_t|$, (2) is by $|\mu_t| \in \{0,1\}$, and (3) is by the calculation in (9.50). Thus, in the i.i.d. case $\operatorname{Var}(G_T(z)) = o(1)$ follows from (9.69) and $\tau T \to \infty$. Also, for all s and some positive constants K_1, K_2, K_3, K_4,

$$|\operatorname{Cov}(\lambda_1,\lambda_{1+s})| \le K_1(\alpha_s^{1-\gamma}[E|\lambda_1|^r]^{1/r}[E|\lambda_1|^p]^{1/p}) \tag{9.70}$$

$$\le K_2(\alpha_s^{1-\gamma}[E|\mu_1|^r]^{1/r}[E|\mu_1|^p]^{1/p}) \tag{9.71}$$

$$\le K_3(\alpha_s^{1-\gamma}[E|\mu_1|]^{\gamma}) \tag{9.72}$$

$$\le K_4\left(\alpha_s^{1-\gamma}\left(\frac{\tau}{T}\right)^{\gamma/2}\right), \tag{9.73}$$

where $1/p + 1/r = \gamma \in (0, 1)$, $p \geq 1$. Here (9.70) follows by Ibragimov's mixing inequality [e.g., Davidson (1994)], (9.71) follows by the previous bound $|\lambda_t| \leq K_0|\mu_t|$ and (9.72) follows by $|\mu_t| \in \{0, 1\}$, while (9.73) follows by (9.69). So $\mathrm{Var}(G_T(z)) = o(1)$ by the condition on the mixing coefficients. $\square$

9.7. *Proof of Theorem* 6.1. Theorem 6.1 is a direct corollary of Theorem 5.1 and Lemma 9.1. Proof of claim (i) follows similarly to the proof in (9.66). Claim (i) implies claims (ii)–(iv), using the properties (v) and (vi) in Lemma 9.1. Uniformity in x in claim (iv) follows from the linearity of $\hat{\rho}_{x,\overline{X},1}$ in x. Finally, claim (v) follows from Theorem 5.1 by the delta method.

9.8. *Tightness of* $Z_\infty(k)$. This section provides primitive conditions for tightness of $Z_\infty(k)$, which is assumed in the statement of Theorem 4.1 and the conditions of uniqueness given in Remark 4.4.

We impose the design condition of Portnoy and Jurečková (1999), who used it for the case $\tau T \to 0$ and show its plausibility on page 233, for example, when $EXX' > 0$. Their proof of tightness is not applicable here, so we have it.

CONDITION PJ. Let F_X denote the distribution function of X. There are a finite integer I, a collection of sets $\{R_1, \dots, R_I\}$ and positive constants δ and η such that:

(a) for each $u \in \{u : \|u\| \geq 1, u_1 \geq 0\}$, there is $R_{j(u)}$ such that $x'u > \delta\|u\|$ for all $x \in R_{j(u)}$,

(b) $\int_{R_j} dF_X > \eta > 0$ for all $j = 1, \dots, I$.

LEMMA 9.7. *If Conditions* R1, R2 *and* PJ *hold, then* $Z_\infty(k)$ *is finite a.s.*

PROOF. Choose $z^f = (z_1^f, \dots, z_d^f)' \in \mathbb{R}^d$ such that

$$Q_\infty(z^f, k) \equiv -k\mu_X' z^f + \int_E (x'z^f - u)^+ \, d\mathbf{N}(u, x) = O_p(1), \tag{9.74}$$

which is possible, as shown in the proof of Theorem 4.1.

Consider a closed ball $B(M)$ with radius M and center z^f, and let $z(k) = z^f + \delta(k)v(k)$, where $v(k) = (v_1(k), \dots, v_d(k))'$ is a direction vector with unity norm $\|v(k)\| = 1$ and $\delta(k) \geq M$. By convexity in z,

$$\frac{M}{\delta(k)}\big(Q_\infty\big(z(k), k\big) - Q_\infty(z^f, k)\big) \geq Q_\infty\big(z^*(k), k\big) - Q_\infty(z^f, k), \tag{9.75}$$

where $z^*(k)$ is a point of boundary of $B(M)$ on the line connecting $z(k)$ and z^f. We will show that, for any K and $\varepsilon > 0$, there is large M such that

$$P\Big(\inf_{v(k) : \|v(k)\|=1} Q_\infty\big(z^*(k), k\big) > K\Big) \geq 1 - \varepsilon. \tag{9.76}$$

(9.76) and (9.74) imply (9.75) $> C > 0$ with probability arbitrarily close to 1 for M sufficiently large, meaning that $Z_\infty(k) \in B(M)$ with probability arbitrarily close to 1 for M sufficiently large, that is, $Z_\infty(k) = O_p(1)$.

Thus, it remains to show (9.76). Since $\mu_X = (1, 0, \dots, 0)'$, $\mu_X' z^*(k) = z_1^f + v_1(k) \cdot M$. Hence, it suffices to show that, for any $\varepsilon > 0$ and any large $K > 0$,

$$-v_1(k) \cdot k \cdot M + \int_E \big(x'z^*(k) - u\big)^+ \, d\mathbf{N}(u, x) \geq K \tag{9.77}$$
$$\text{w.pr.} \geq 1 - \varepsilon, \text{ for large enough } M,$$

and, therefore, we establish (9.76). We have by Condition PJ that, for some $R_{j(v)}$ with $j(v) \in \{1, \dots, I\}$,

$$\begin{aligned} &\int_E \big(x'z^*(k) - u\big)^+ \, d\mathbf{N}(u, x) \\ &\quad \geq \int_{([-\infty,\kappa]\times R_{j(v)})\cap E} \big(x'z^*(k) - u\big)^+ \, d\mathbf{N}(u, x) \\ &\quad \geq \mathbf{N}\big(([-\infty, \kappa] \times R_{j(v)}) \cap E\big) \times (\delta M - \kappa - \kappa')^+, \end{aligned} \tag{9.78}$$

where $\kappa \in \mathbb{R}$ is a constant to be determined later and that does not depend on $v(k)$ and $\kappa' = \max_{x \in \mathbf{X}} |x'z^f|$.

Note that for any region $\mathbb{X}$ such that $\int_{\mathbb{X}} dF_X > \eta > 0$ and any $\kappa_1 > 0$ and $\varepsilon > 0$, there is a sufficiently large κ_2 such that

$$\mathbf{N}\big(([-\infty, \kappa_2] \times \mathbb{X}) \cap E\big) > \kappa_1 \qquad \text{w.pr.} \geq 1 - \varepsilon. \tag{9.79}$$

Hence, by (9.79) we can select κ large enough so that

$$\mathbf{N}\big(([-\infty, \kappa] \times R_j) \cap E\big) > \frac{(k+1)}{\delta} \tag{9.80}$$
$$\text{for all } j \in \{1, \dots, I\} \text{ w.pr.} \geq 1 - \varepsilon,$$

so that w.pr. $\geq 1 - \varepsilon$,

$$\begin{aligned} &-v_1(k) \cdot k \cdot M + \int_E (x'z^*(k) - u)^+ \, d\mathbf{N}(u, x) \\ &\quad \geq -k \cdot M + (k+1) \frac{(\delta M - \kappa - \kappa')^+}{\delta}. \end{aligned} \tag{9.81}$$

Now set M sufficiently large to obtain (9.77). $\square$

Acknowledgments. This paper is based on a chapter of my Stanford Ph.D. dissertation completed in August 2000, and I would like to thank my dissertation committee: Takeshi Amemiya, Pat Bajari and Tom Macurdy. I would also like to thank Roger Koenker, Stephen Portnoy, Ivan Fernandez–Val, Jerry Hausman, Bo Honore, Jana Jurečková and Keith Knight. I am also very grateful to two anonymous referees, an Associate Editor and Co-Editors of the journal for providing feedback of the highest quality.

REFERENCES

ABREVAYA, J. (2001). The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics* **26** 247–257.

BUCHINSKY, M. (1994). Changes in the U.S. wage structure 1963–1987: Application of quantile regression. *Econometrica* **62** 405–458.

CADE, B. (2003). Quantile regression models of animal habitat relastionships. Ph.D. dissertation, Colorado State Univ. Available at www.fort.usgs.gov.

CHAMBERLAIN, G. (1994). Quantile regression, censoring, and the structure of wages. In *Advances in Econometrics*: *Sixth World Congress* (C. Sims, ed.). Cambridge Univ. Press.

CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777.

CHAUDHURI, P., DOKSUM, K. and SAMAROV, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25** 715–744.

CHERNOZHUKOV, V. (1998). Nonparametric extreme regression quantiles. Working paper, Stanford Univ. Presented at Princeton Econometrics Seminar, December 1998.

CHERNOZHUKOV, V. (1999). Conditional extremes and near-extremes: Estimation, inference, and economic applications. Ph.D. dissertation, Dept. Economics, Stanford Univ. Available at www.mit.edu/~vchern.

CHERNOZHUKOV, V. and UMANTSEV, L. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* **26** 271–292.

DAVIDSON, J. (1994). *Stochastic Limit Theory*. Oxford Univ. Press, New York.

DE HAAN, L. (1984). Slow variation and characterization of domains of attraction. In *Statistical Extremes and Applications* (I. Tiago de Oliveira, ed.) 31–48. Reidel, Dordrecht.

DE HAAN, L. and ROOTZÉN, H. (1993). On the estimation of high quantiles. *J. Statist. Plann. Inference* **35** 1–13.

DEKKERS, A. and DE HAAN, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** 1795–1832.

DOKSUM, K. A. and GASKO, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *Internat. Statist. Rev.* **58** 243–252.

DONALD, S. G. and PAARSCH, H. J. (1993). Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *Internat. Econom. Rev.* **34** 121–148.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events*. Springer, Berlin.

FEIGIN, P. D. and RESNICK, S. I. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51** 135–165.

GEYER, C. J. (1996). On the asymptotics of convex stochastic optimization. Technical report, Dept. Statistics, Univ. Minnesota.

GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20** 305–330.

HE, X. (1997). Quantile curves without crossing. *Amer. Statist.* **51** 186–192.

HENDRICKS, W. and KOENKER, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Amer. Statist. Assoc.* **87** 58–68.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26** 755–770.

KNIGHT, K. (1999). Epi-convergence and stochastic equisemicontinuity. Technical report, Dept. Statistics, Univ. Toronto, Available at www.utstat.toronto.edu/keith/papers/.

KNIGHT, K. (2001). Limiting distributions of linear programming estimators. *Extremes* **4** 87–103.

KOENKER, R. and BASSETT, G. S. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOENKER, R. and BASSETT, G. S. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50** 43–61.

KOENKER, R. and GELING, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Amer. Statist. Assoc.* **96** 458–468.

KOENKER, R. and PORTNOY, S. (1987). L-estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

KOROSTELËV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489.

LAPLACE, P.-S. (1818). *Théorie Analytique des Probabilités*. Éditions Jacques Gabay (1995), Paris.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

MEYER, R. M. (1973). A Poisson-type limit theorem for mixing sequences of dependent "rare" events. *Ann. Probab.* **1** 480–483.

PICKANDS, III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131.

PORTNOY, S. (1991a). Asymptotic behavior of regression quantiles in nonstationary, dependent cases. *J. Multivariate Anal.* **38** 100–113.

PORTNOY, S. (1991b). Asymptotic behavior of the number of regression quantile breakpoints. *SIAM J. Sci. Statist. Comput.* **12** 867–883.

PORTNOY, S. and JUREČKOVÁ, J. (1999). On extreme regression quantiles. *Extremes* **2** 227–243.

PORTNOY, S. and KOENKER, R. (1997). The Gaussian hare and the Laplacian tortoise (with discussion). *Statist. Sci.* **12** 279–300.

POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

SMITH, R. L. (1994). Nonregular regression. *Biometrika* **81** 173–183.

TSAY, R. S. (2002). *Analysis of Financial Time Series*. Wiley, New York.

WATTS, V., ROOTZÉN, H. and LEADBETTER, M. R. (1982). On limiting distributions of intermediate order statistics from stationary sequences. *Ann. Probab.* **10** 653–662.

DEPARTMENT OF ECONOMICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02142
USA
E-MAIL: vchern@MIT.edu