Massachusetts Institute of Technology
Department of Economics
Working Paper Series

CONDITIONAL EXTREMES AND NEAR-EXTREMES

Victor Chernozhukov, MIT

Working Paper 01-21
July 2000

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

CONDITIONAL EXTREMES AND NEAR-EXTREMES

Victor Chernozhukov, MIT

Working Paper 01-21
July 2000

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

CONDITIONAL EXTREMES AND NEAR-EXTREMES

Victor Chernozhukov[1]
Dept. of Economics, MIT, E52-262F,
Cambridge, MA, 02142
email: vchern@mit.edu

Version: May 10, 2001
First Version: October, 1998

Abstract

This paper develops a theory of *high* and *low* (extremal) quantile regression: the linear models, estimation, and inference. In particular, the models coherently combine the convenient, flexible linearity with the extreme-value-theoretic restrictions on tails and the general heteroscedasticity forms. Within these models, the limit laws for extremal quantile regression statistics are obtained under the rank conditions (experiments) constructed to reflect the extremal or rare nature of tail events. An inference framework is discussed. The results apply to cross-section (and possibly dependent) data. The applications, ranging from the analysis of babies' very low birthweights, (S, s) models, tail analysis in heteroscedastic regression models, outlier-robust inference in auction models, and decision-making under extreme uncertainty, provide the motivation and applications of this theory.

Key Words: Quantile Regression

 This paper is a part of my Ph.D. dissertation at Stanford, completed in 2000. I am very grateful to Takeshi Amemiya, Roger Koenker, Thomas MaCurdy, Mordecai Kurz, and also to Anil Bera, Stephen Portnoy, Thomas Sargent for advise and help; Pat Bajari, Joseph Romano, Keith Knight, and Jana Jureckova, for discussions and from whom I learned a great deal; and also friends: Igor Makarov, Stijn Van Nieuwerburgh, Leonardo Rezende, Michael Schwarz, and Lei Zhang, and especially Han Hong, Robert McMillan, Sergei Morozov, and Len Umancev. I thank also the seminar participants at Brown, Caltech, Chicago, CIRANO, Duke, Mannheim, MIT-Harvard, Northwestern, Stanford, UCLA, UCSD, University of Illinois, Yale, and Neuchatel May 2000 Conference on Recent Developments in Regression, CEME 2001 Conference, Conference on Economic Applications of Quantile Regression in Konstanz, 2000, especially J. Angrist, D.Andrews, J. Altonji, M. Buchinsky, P. Christophersen, F. Diebold, G. Elliot, R. Engle, B.Fitzenberger, C. Granger, J. Hahn, J. Hamilton, L. Hansen, M. Hallin, J.Heckman, G. Imbens, D. Jorgenson, T.Bollerslev, C. Manski, R. Matzkin, P. Phillips, J. Rust, R. Sherman, D. Strook, G. Tauchen, R. Tsay, J. Jureckova, and H. White. I also thank Jerry Hausman, Guido Kursteiner, Ziad Nejmeldeen, and Whitney Newey for suggestions that drastically improved the presentation of the material; and also Bo Honore and Charles Manski for suggesting to look into modeling questions carefully.

I gratefully acknowledge the financial support provided by the Alfred P. Sloan Dissertation Fellowship.

Contents

1 Introduction

Regression quantiles, Koenker and Bassett[48], represent a flexible and informative method of regression analysis as they describe the conditional distribution of the response variable Y given covariate X, without imposing rigid distributional assumptions. The goal of this paper is to model and make inference on the *extremal* (*near-extreme high or low*) regression quantile functions. In essence, they represent the models of the extremal values of Y conditional upon X. For example, the near-extreme 0.1-th conditional quantile function describes the values below which Y falls with probability 10% given values of X.

Modeling high or low conditional quantiles is motivated by many examples. Some include: (i) in *micro-economics*: (S, s) models of investment, inventory, employment shortages; auction models, reservation wage equations; (ii) in *finance*, *micro-* and *macro-economics*: decision-making under extreme uncertainty, where good risk measures are vital for the purposes of insurance, safety-first resource allocation, and management of risks; and many others.

The ordinary extremal quantiles, the models and the sample analogs, have been the main subject of classical and modern extreme value theory, which forms an important field of applied and theoretical statistics.[2] The theory was developed by Von Mises, Frechet, Fisher, Gnedenko, Smirnov, de Haan, and many others. The ordinary sample quantiles have an immense inference role, providing the estimators of the tail index and other tail functionals (Pickands[57], Hill[40], Dekkers and de Haan[21]). Analogous motivations underlie the present analysis as well.

In this paper we study the extremal (high and low) conditional quantiles – the linear *models and* the sample regression *analogs.* In particular, the models coherently combine convenient, flexible linearity with the extreme-value-theoretic restrictions on tails and the general heteroscedasticity forms. Within these models, the limit laws for extremal quantile regression statistics are obtained under the rank conditions constructed to reflect the extremal or rare nature of tail events. The *goal is* the practical, important problem of *modeling* and *making inference* on the *.3-th and lower and .7-th and higher* regression quantiles, as well as conducting the *tail inference*, in the common economic data sets.[3] (*Our target is not* the "exotic" 0-th quantile.)

The *rank conditions* approximate the degrees of *lack of data* or *extremality* pertinent to the inference about the quantiles of interest. Define rank r as the quantile index τ times the sample size T. The extreme and intermediate rank conditions apply to the cases where index τ is *extremal* (e.g. .1, .2) and is *low* (r is small) or *not low* (r is large) relative to the sample size T.[4] Formally, the sequence of the quantile index-sample size pairs (τ_T, T) is an *extreme rank sequence* if

$$(i) \quad \tau_T \searrow 0, \quad \tau_T T \to k > 0,$$

and an *intermediate rank sequence* if

$$(ii) \quad \tau_T \searrow 0, \quad \tau_T T \to \infty.$$

[2] As of today, thousands of papers are devoted to the extreme value theory. Many excellent books give systematic treatments. See e.g. [7], [64], [65], [52],[33],[34],[26], [73].

[3] Say with $T \leq 3000$, and typical number of regressors, $5 - 10$ and higher.

[4] Heuristically, r is number of observations to make inference on the τ- regression quantile.

Because the principles (i) and (ii) *constructively exploit* that the *relevant data is formed by the tail events and, or is scarce*, they lead to

a. asymptotic distributions that either fit the finite-sample distributions better or are more parsimonious than the conventional approximations,

b. important tail inferences, based on the extremal regression quantiles.

In evaluating these concepts, it is important to keep in mind that these *alternative sequences* are designed to *yield better approximations* in given practical problems with given sample sizes, even when the quantile index is not very low. And, in case(a), it is completely irrelevant whether or not *future* sampling will lead to samples conforming to these sequences or not.

The concepts (i) and (ii) are well motivated by the intellectual and practical success of the extreme value theory, which focused on the ordinary sample quantiles. The concepts are also similar in spirit to other "alternative" asymptotics, e.g., GMM when the number of moment conditions is large; weak instruments theory, where the instruments are weakly correlated with the regressors; near-to-unit root theory; and, generally, the theory of statistical experiments.

The organization and contribution of this paper is as follows:

1. Section 2 demonstrates the relevance of the problem in economic analysis.

2. Section 3 introduces the principles of extremality for the regression quantiles – namely that of the intermediate and extreme rank sequences.

3. Section 4 develops the models of the extremal (low) conditional quantiles. They *coherently* combine the linear functional forms with the extreme-value-theoretic restrictions, and lead to non-degenerate, parsimonious limit distributions. The models are distribution-free and flexible, allowing for sophisticated effects of covariates on the shape of the conditional distribution (the scale, kurtosis, skewness, etc.). Importantly, these models *do not admit reductions* to the classical one-sample case (by removing the conditional mean and/or scale).

4. Within the formulated models, section 5 provides the asymptotic limit theory for the sample regression quantiles under the extreme rank condition, $\tau T \to k > 0$.[5] The limit is driven by a stochastic *integral* of a "residual" function with respect to a Poisson point process.

5. Section 6, using additional tail restrictions, provides the asymptotic distributions of regression quantiles under the intermediate rank condition, $\tau T \to \infty, \tau \to 0$. The limit is normal, with variance parsimoniously determined by the tail indices. This enables a very practical inference. (In contrast, the conventional theory requires the nonparametric estimates of the conditional density functions evaluated at the extremal quantiles). This provides a regression analogue of fairly recent results of Dekkers and de Haan[21].

[5]This paper is not about $\tau = 0$, the linear programming estimator (also called 'extreme regression quantile'), considered in Feigin and Resnick[29], Portnoy and Jureckova[60], Chernozhukov[12], Knight[47] within the location-shift model (Covariates only affect the location but not the scale, shape or tail of the conditional distribution.) The estimator, defined as $\max \bar{X}'\beta$ s.t. $Y_t \leq X_t\beta, \forall t$ and useful as a boundary estimate, can't be used at all in the present context. We look at different estimators (high and low regression quantiles) that have very different asymptotics and applications [$\tau > 0$ (T is finite); see examples 2.1-2.4, where the support is unbounded or "boundaries" depend on unobserved variables]. We also develop and operate with very different models.

6. We conclude by discussing an inference theory and an empirical paper [15].

Also relevant are the works of Smith[73], Tsay[75], and references therein, who develop the models of exceedances over high constant thresholds. The parametric likelihood of the Pareto family is used to describe such data, and parameters are made dependent on regressors. It should be clear that the goals, models, and methods of this paper are quite different. Our analysis should be viewed as complementing the study of the central rank regression quantiles, with the motivation stemming from the wide use of quantile regression in data analysis in econometrics and statistics.

2 Econometric Applications

Quantile regression is a popular tool in econometric applications. See Abadie et al.[1], Buchinsky[9], Chamberlain[11], Poterba and Rubin[61], and the review of Koenker and Hallock[50]. Our results can be useful in about any such application, since our focus is the inference about high and low conditional quantiles (say .7 and higher and .3 and lower, in a typical data-set), recognizing the extremality and/or scarcity of the tail events. Important inferences about the tail shapes can be made as well. There are many examples, where high or low quantiles are of particular interest. For example, Abreveya[2] and Koenker and Hallock[50] characterize the economic determinants of babies' very low birth-weights through the near-extreme conditional quantiles (.05 and below). Deaton[20] examines food expenditure of Pakistani households by the .1-th and .9 -th conditional quantiles. The following presents a brief discussion of some others.

Example 2.1 (Determinants of Generalized (S, s) Models.)

The (S, s) theory is widely used in the firm-level *microeconomic* studies, including the analysis of durable good inventories, employment shortages, and investment in capital goods. Lumpiness of adjustments, a main prediction, is well documented. E.g., Arrow et al. [5], Scarf[71], Rust and Hall[37], Aguirregabiria[3], Caballero [10].

In the (S, s) theory, a firm allows a state variable Y_t (capital stock, inventory) to fall until it reaches a lower barrier, $s(X_t)$, at which point the stock is replenished (jumps) to an upper barrier, $S(X_t)$. Such decisions are optimal in general settings (Hall and Rust[38]). X_t may include prices and other variables that affect the firm's beliefs about future sales and costs (e.g. industrial production and commodity price indices, interest rates). Assume that (Y_t, X_t) are observed for a cross-section of firms. Absent unobserved heterogeneity, $s(X_t)$ and $S(X_t)$ are exactly the minimal and maximal conditional quantiles of Y_t, given X_t. Otherwise, $s(X_t)$ and $S(X_t)$ are still strongly related to the extremal conditional quantiles.

To address the *unobserved heterogeneity*, Caballero and Engel[10] introduce the stochastic barriers $[s(X_t) - v_t, S(X_t) + e_t]$ with unobserved time- and firm- specific random components e_t, v_t. They propose probabilistic adjustment models (hazard functions) to describe the evolution of Y_t across firms and or times. In such models, the high and low conditional quantiles also describe the probabilistic rules. For example, the inventory variable Y_t is below the .1-th conditional quantile only with probability 10%, given X_t. Inference about such functions is exactly our area of focus. More generally, we can map quantile functions into hazard functions, and vice-versa.

Apart from descriptive analysis, extremal quantile regression can estimate the de-

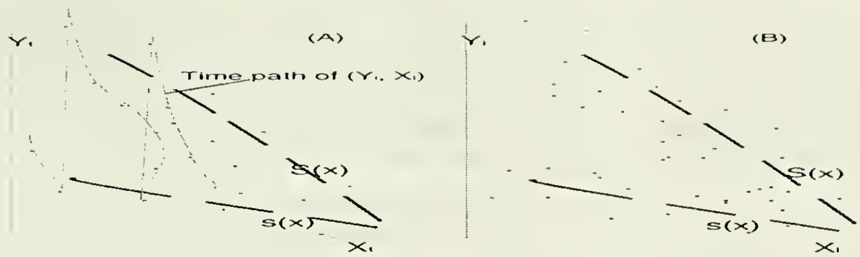

Figure 1: (S, s) model with stochastic bands $[s(X) - v, S(x) + e]$, where e and v are the unobserved firm and time specific random components. Panel (A): Data on a single firm may be generated by discrete sampling from the time path of (Y_t, X_t). E.g. Rust and Hall[37]. Panel (B): Data (Y_i, X_i) may be generated as a cross-section of plants. E.g. [3],[10].

terminants of the $(S(x), s(x))$ functions. Specifically, suppose that

$$[s(X_t) - v_t, S(X_t) + e_t]$$

constitute the adjustment barriers, with $v_t, e_t > 0, s(X_t) < S(X_t)$ a.s., so that the interval is non-empty. The timing is continuous. If Y_t hits the lower bound $s(X_t) - v_t$, it is adjusted to the upper bound $S(X_t) + e_t$. Pairs (Y_t, X_t) are the observed draws of different firms (or a panel, stationarity assumed). For brevity, let's focus on $s(X)$.

Suppose v_t and e_t are independent of X_t,[6] then for $c \geq 0$:

$$\begin{aligned} P(Y_t < s(X_t) - c|X_t) = &EP(Y_t - s(X_t) < -c|X_t, c < v_t) \cdot P(c < v_t) \\ &+ EP(Y_t - s(X_t) < -c|X_t, c > v_t) \cdot P(c > v_t). \end{aligned}$$

By construction $P(Y_t - s(X_t) < -c|X_t, c > v_t) = 0$. Additionally, impose the following tail homogeneity condition: for all $c > 0$ sufficiently close to v_t:[7]

$$P(Y_t - s(X_t) < -c|X_t, c < v_t) = \alpha(v_t - c). \tag{2.1}$$

Thinking of $Y_t - [s(X_t) - v_t]$ as a positive "duration" variable, (2.1) states an "accelerated failure time" model for the tail (which is more general than in [10]). (2.1) imposes no restrictions on the central features of the conditional distribution of Y_t, which is reasonable, since the (S, s) theory does not relate the central features to the adjustment barriers. For example, the symmetry or homoscedasticity assumptions are unreasonable. This implies that for $-c$ low enough and some low constant $\phi(c)$

$$P(Y_t - s(X_t) < -c|X_t) = \phi(c),$$

or, equivalently, that for small $\tau > 0$

$$Q_{Y_t}(\tau|X_t = x) \equiv s(x) - c(\tau)$$

[6]This is reasonable, since $S(X), s(x)$ incorporate the barrier component that depends on X. Nevertheless, we can allow e, v to be dependent on X. A note is available upon request.

[7]$=$ can replaced by $\sim$, as c increases.

is the τ-th *conditional quantile* of Y_t given X_t. Therefore, $s(x)$ equals the low (extremal) conditional quantiles up to an additive constant. Notably, it is not possible to estimate $s(x)$ off the central features of the conditional distribution of Y_t, as discussed above. The inference about $Q_Y(\tau|X)$ for low values of τ is exactly our area of focus. The analytical examples of Rust and Hall suggest that linear/polynomial functions are excellent descriptions of $(s(x), S(x))$ functions.

Example 2.2 (Tail Analysis in Regression Models) The tail shape (index) of the conditional distribution is important in the regression analysis. For example, the thick-tailed distributions favor the LAD and other estimators more than the OLS. Thus, knowing the *tail index* helps determine better estimators. On the other hand, the tail shapes are important in describing the large insurance claims ([26]), the analysis of the long and short term survival and durations ([49], [42]), and financial data (e.g. Mandelbrot[54], Fama[28], Kearns and Pagan[45], Danielsson and de Vries[17]). In the non-regression setting, the tail index estimators of Hill and Pickands have been countlessly used in the empirical analysis. However, estimation of the tail index *in the presence of the shape heteroscedasticity* (scale, skewness, kurtosis, and other forms) is largely an open, difficult problem. Our results allow one to construct the *regression analogs* of the Pickands and Hill tail index estimators, based on the extremal regression quantiles, which specifically *adapt* to the *shape-heteroscedastic* setting, and are simple in practice. Section 7 offers a discussion, and [15] provides an empirical application.

Example 2.3 (Decision Making under Extreme Uncertainty) Risk is a key subject of non-financial and financial decisions, insurance, and regulation. Both the firms and the regulators are seriously concerned about extreme risks – the tail events that can wipe out capital, hindering liquidity or solvency.

An important branch of economics literature is devoted to safety-first decision making. See Roy[68], Telser[74], Pyle and Turnovsky[63], Bertail et al[8] and others. In this approach, the decision-makers (firms, investors, regulators) solve either:

$$1. \max_{\alpha:\ Q_{Y_t(\alpha)}(\tau|X_t)\leq \bar{z}} \bar{z}, \qquad \text{or} \quad 2. \max_{\alpha:\ Q_{Y_t(\alpha)}(\tau|X_t)\leq \bar{z}} \mu(\alpha),$$

where $Y_t(\alpha)$ is the random payoff (e.g. private or public benefits and profits) to the decision α (technology, portfolio composition, buffer stocks, quality/quantity of food control); $\bar{z}$ is the *safety margin* or disaster level of the payoff; τ is the *probability of the disaster* or of exceeding the margin, set to be *small*; μ is the mean of $Y_t(\alpha)$; $Q_{Y_t(\alpha)}(\tau|X_t)$ is the conditional τ-th quantile function of $Y_t(\alpha)$ given X_t, the vector of variables representing the current state. $Q_{Y_t(\alpha)}(\tau|X_t) \leq \bar{z}$ is the conditional (extremal) quantile constraint, requiring the disaster probability to be small: $P(Y_t(\alpha) \leq \bar{z}|X_t) \leq \tau$. This presents a problem of inference concerning the conditional extremal quantiles. Our models are flexible (central features of the distribution do not determine the tail features) and *specifically* exploit the extremality and scarcity of the tail events.

In Chernozhukov and Umantsev [15], we apply the present results.

Safety-first decisions are very important in the finance industry, where quantiles (value-at-risk) are the *required* measures of the high level infrequent risk, used to determine the capital requirements and other external and internal purposes. See [25], [56], [27], [35], among others, for a sample of illuminating research as well as reviews. Value-at-risk is computed as the level below which the (daily or weakly) return is only

1% or 5% of the time (.01-th and .05-th quantiles). Again, this is a problem concerning the conditional extremal quantiles.

Example 2.4 (Simple Robust Inference in Boundary-Dependent Models) Parametric boundary dependent likelihoods, arising in the models of job search and auctions (see [16], [31], [23], [41], [36] for a sample of remarkable works) take the form:

$$L(\beta,\gamma) = \sum_t \ln f(Y_t|X_t,\gamma,\beta)\cdot 1(Y_t \geq X_t'\beta),$$

where $f(X_t'\beta|X_t,\gamma,\beta) > 0$ a.s. and is finite. β and γ are the boundary and shape parameters, respectively. Linearity of the boundary is not essential (see below).

Likelihood procedures, e.g. ML, estimate γ and β jointly. The estimates $\hat{\beta}$ are characterized by $d \equiv \dim(X)$ constraints, $Y_t = X_t'\hat{\beta}$, where Y_t is among the extremal values of Y_t, [23] and[41]. For example, $\min_{t\leq T} Y_t$ is the boundary estimate in the no-regressor case. Therefore, having a *few outlier* observations Y_t^o (such that $Y_t^o < X_t'\beta$,) *severely biases* and renders inconsistent the estimates of both β and γ. The outliers arise as misrecordings of the bid with a low probability (not the usual additive measurement error) or bid mistakes. Bajari[6] offers a substantive analysis, suggesting outliers are responsible for drastic overestimates of the mark-ups in prominent auction studies.

Suppose the number of outliers Y_t^o is bounded by a constant K, independent of T. Consider the τ-th near-extreme regression quantile estimator $x'\hat{\beta}(\tau)$ of the boundary $x'\beta$, with quantile index $\tau = M/T$, $M \sim \ln T$. Asymptotically $x \mapsto x'\hat{\beta}(\tau)$ passes above the outliers, and is $T/\ln T$-rate-consistent. Substitute $\hat{\beta}(\tau)$ into $L(\beta,\gamma)$ and estimate γ via ML. The resulting estimator of γ is efficient. Chernozhukov and Hong[14] offer an analysis. Although we focus on the linear boundaries, a non-linear extension in this model is straightforward. Regardlessly, the linear forms include the polynomial and piece-wise linear specifications, approximating the smooth parametric functions as well as we like.

3 Extremal Quantiles and Rank Sequences

This section defines the linear regression model, the sample regression quantiles, the extreme and intermediate rank concepts for these statistics, and the tail types.

3.1 Extremal Conditional Quantiles

Suppose Y_t is the response variable in $\mathbb{R}$, and X_t are the conditioning variables in $\mathbb{R}^d$. The τ-th conditional quantile function $Q_Y(\tau|x)$ is a function $q(x)$ that satisfies the relationship $P(Y \leq q(X)|X) = \tau$. For instance, $Q_Y(.25|x)$ and $Q_Y(.1|x)$ are the conditional first quartile and decile functions. Formally,

$$Q_Y(\tau|x) \equiv F_Y^{-1}(\tau|x),$$

where $F_Y^{-1}(\cdot|x)$ is the inverse of $F_Y(\cdot|x)$. Our focus is exclusively on modeling and making inference on the extremal conditional quantile functions:

$$Q_Y(\tau|x), \text{ where } \tau \text{ is } \textit{near } 0.$$

The formal concept of *extremality* or *nearness* will be developed later.

3.2 Linear Quantile Regression Model

In this paper we consider the linear model for quantiles of interest $\mathcal{I}$

$$Q_Y(\tau|x) = F_Y^{-1}(\tau|x) = x'\beta(\tau), \quad \forall \tau \in \mathcal{I}, \tag{3.2}$$

where $\beta(\cdot)$ is an *unknown function* of τ. Here it is necessary that (3.2) holds for

$$\mathcal{I} = [0, \eta], \text{ where } \eta > 0. \tag{3.3}$$

If η is small, the linearity is assumed only for low quantiles, and not necessarily for other quantiles. We assume that X has (or is trimmed to) a *compact* support $\mathbf{X}$.

The model (3.2) is implied, for instance, by the classical linear *location-scale* models with *unknown* error distribution, but is considerably more flexible in the sense that the *shape* of the conditional density may change with the covariates. X may incorporate a wide array of polynomial and other transformations of the observed covariates. On its basis, in section 4, we develop the models with the extreme-value-theoretic restrictions on the conditional tails.

Note the approaches to linear modeling. One approach assumes linearity of a single or few quantiles (Buchinsky[9], Horowitz[43], Powell[62]). Another approach (Koenker and Machado[51]) assumes the linearity of all quantile functions, $\mathcal{I} = [0, 1]$. The "local in τ linearity" assumption made here is closer to the first approach.

Despite convenience, having linearity for several τ may pose an avoidable caveat (the curves may cross). First, X is often a transformation of the original covariates, so the curves are non-linear in the original space (see [49]). Second, given compactness of support $\mathbf{X}$, the linear model is always coherent. Take a countable, possibly finite collection of non-crossing curves $\{x \mapsto x'\beta(\tau_i), i \in J\}$ with domain $\mathbf{X}$. Define $x \mapsto x'\beta(\tau)$ for other τ by taking appropriate convex or linear combinations of these lines. By construction, the lines cross only outside $\mathbf{X}$. This also defines the conditional c.d.f.

3.3 Sample Regression Quantile Statistics

Suppose we have T observations $\{Y_t, X_t\}$. In the no-covariates case the sample τ-th quantile $\hat{\beta}(\tau)$, is generated by solving the problem

$$\min_{\beta \in \mathbb{R}} \sum_{t=1}^{T} \rho_\tau(Y_t - \beta),$$

where $\rho_\tau(x) \equiv (\tau - 1(x \leq 0))\, x$. Koenker and Bassett[48] extended the concept to the regression setting by solving

$$\min_{\beta \in \mathbb{R}^d} \sum_{t=1}^{T} \rho_\tau\left(Y_t - X_t'\beta\right). \tag{3.4}$$

The $\hat{\beta}(\tau)$ that solves (3.4) has the equivariance and robustness properties of the ordinary sample quantiles; in particular, **(i)** regression equivariance, **(ii)** scale equivariance, **(iii)** equivariance to (full rank) linear transformations of X, **(iv)** invariance to perturbations of Y_t without crossing the hyperplane $x'\hat{\beta}(\tau)$. The solutions $x'\hat{\beta}(\tau)$ to (3.4) (if unique) pass through d points (Y_t, X_t) and the function $\tau \mapsto \bar{X}'\hat{\beta}(\tau)$ is monotone in τ.

3.4 Extremes, Near-Extremes & Data Scarcity

We view the sample regression quantiles as order statistics in regression settings. For a given sample of size T, the τ-th sample regression quantile is seen here as the τT-th order statistic. Henceforth, we shall refer to τT as to the *rank* or *order*.

Definition 3.1 (Rank Conditions) The sequence of quantile index-sample size pairs (τ_T, T) is said to be:

(i) an extreme rank sequence, if $\tau_T \searrow 0$, $\tau_T T \to k > 0$,

(ii) an intermediate rank sequence, if $\tau_T \searrow 0$, $\tau_T T \to \infty$,

(ii) a central rank sequence, if τ is fixed, and $T \to \infty$.

Even though (i) and (ii) make τ sample size dependent, to simplify we write τ instead of τ_T. Because principles (i) and (ii) constructively exploit that the *relevant data is formed by the tail events and, or is scarce*, they lead to

a. asymptotic distributions that either fit the finite-sample distribution better or, given the same approximation quality, are more parsimonious relative to the conventional central rank approximations (Koenker and Bassett[48], Powell[62])

b. important tail inference procedures, based on the sample regression quantiles. See example 2.3 and section 7.

Concepts (i) and (ii) are well motivated by the intellectual and practical success of extreme value theory, which focused on the ordinary sample quantiles. These concepts are also similar in spirit to other types of "alternative" asymptotics, e.g. GMM when the number of moment conditions is large, or, generally, the theory of statistical experiments.

In evaluating these concepts, it is important to keep in mind that these alternative sequences are designed to yield practically better approximations even when the quantile index is not very low. And, in case (a), it is completely irrelevant whether or not *future* sampling will lead to samples conforming to these sequences.

To clarify (a), consider a simple example with no X. Suppose, with an i.i.d. sample $\{U_t, t \le T \equiv 200\}$, we wish to infer about the quantiles with indices $\tau = .025$, $\tau = .1$, $\tau = .2$, $\tau = .3$. The estimators are the order statistics (sample quantiles) $U_{(5)}, U_{(20)}, U_{(40)}, U_{(60)}$. Suppose the distribution F_U has an algebraic tail $F_U(x) \sim (-x)^{-1/\xi}, \xi = 1$ as $x \searrow -\infty$. Figure 5 compares the conventional central rank approximation $\sqrt{T}\left(U_{(\tau T)} - F^{-1}(\tau)\right) \xrightarrow{d} N\left(0, \tau(1-\tau)/f_U^2(F_U(\tau))\right)$, where $f_U \equiv F_U'$, with the intermediate rank one: $a_T(U_{(\tau T)} - F^{-1}(\tau)) \xrightarrow{d} N(0, \xi^2/(m^{-\xi} - 1)^2), (\xi = -1, m > 1)$, $a_T = \sqrt{\tau T}/F^{-1}(m\tau) - F^{-1}(\tau))$, and the extreme rank approximation: $T^{-1/\xi}(U_{(\tau T)} - F^{-1}(\tau)) \xrightarrow{d} -k^{-1/\xi} - \Gamma_k^{-1/\xi}$, where Γ_k is a gamma random variable with degree k (sum of k standard exponentials, section 3.6).

Quality-wise, the *extreme rank* approximation, which *exploits both the extremality and scarcity* of tail events, beats the normal quite considerably (displays A-C). Only for a fairly non-extreme quantile, $\tau = .3$, does the normal approximation achieve roughly the same quality. At the same time, the *intermediate rank* approximation, which *exploits the extremality* of relevant events, is very close to the central rank approximation (displays D-F), but enjoys greater parsimony and ease of inference. The tail index ξ is easy to estimate, and the scaling a_T is estimated by the sample interquantile spacing (see section 3.6). This may be preferred to the nonparametric estimation of the density function evaluated at the low quantile (with the scarce tail data), as required in the central rank theory. See [13] for a Monte-Carlo regression example.

3.5 Tail Types, Support Types, and Classical Limits

The following definitions are important in the sequel.

Definition 3.2 (Types of Support) In view of linearity, we say $F_Y(\cdot|X)$ has:

- finite support, if $Q_Y(0|X) > -\infty$, a.s.
- infinite support, if $Q_Y(0|X) = -\infty$, a.s.

Definition 3.3 (Tail Types, Tail Index, Regular variation) Consider a random variable U with distribution function F_u, with lower end-point x_f equal 0 or $-\infty$. F_u has the tail of the *extremal types 1, 2, or 3* if for [$f \sim g$ if $f/g \to 1$.]

$$\begin{aligned}
&\text{type 1: } (\ \xi \equiv 0\ ) &&: \text{ as } t \searrow 0 \text{ or } -\infty, && F_u(t + x\ell(t)) \sim F_u(t)e^{-x}, \quad \forall x\ \in \mathbb{R}, \\
&\text{type 2: } (\xi \equiv \frac{1}{\alpha}) &&: \text{ as } t \searrow -\infty, && x^{-\alpha}F_u(t) \sim F_u(tx),\ \forall x > 0,\ \alpha > 0, \\
&\text{type 3: } (\xi \equiv \frac{-1}{\alpha}) &&: \text{ as } t \searrow 0, && x^{\alpha}\ F_u(t) \sim F_u(tx), \quad \forall x > 0,\ \alpha > 0.
\end{aligned} \tag{3.5}$$

where $\ell(t) = \int_{x_f}^{t} F_u(v)dv/F_u(t)$, for $t > x_f$, cf. [52]. Enclosed in the brackets in (3.5) is the *tail index* ξ, which determines the tail type.

Equation (3.5) defines type 2 distributions as *regularly varying functions at* $-\infty$ with *index* $-1/\xi = -\alpha$, (algebraically and near-algebraically tailed at $-\infty$, in more intuitive terms). (3.5) also defines type 3 distributions as *regularly varying functions at* 0 with *index* $-1/\xi = \alpha > 0$ (algebraically and near-algebraically tailed at a finite end point, taken as 0). The type 1 class includes exponentially and near-exponentially tailed distributions. For future reference, note these conditions imply that the quantile function $F_u^{-1}(\tau)$ is *regularly varying at* 0 *with index* $-\xi$. E.g. [26].

Classes 1-3 contain most of smooth distributions with rare exceptions. See [26]. The tail types determine the limiting distributions of order statistics under extreme and intermediate ranks. In our setting, they will have a similar role as well. For later comparisons, let us review the non-regression results.

3.6 Limit Distributions of Ordinary Sample Quantiles

Extreme Rank Statistics. Consider the order statistics $U_{(1)} \leq ... \leq U_{(k)}$ from the i.i.d. sample $U_1, ..., U_T$, distributed according to law F_u, with the lower end-point x_f equal 0 or $-\infty$. The extreme value theory described the existence and forms of the non-degenerate limit laws for the properly normalized order statistics:

$$a_T(U_{(k)} - b_T).$$

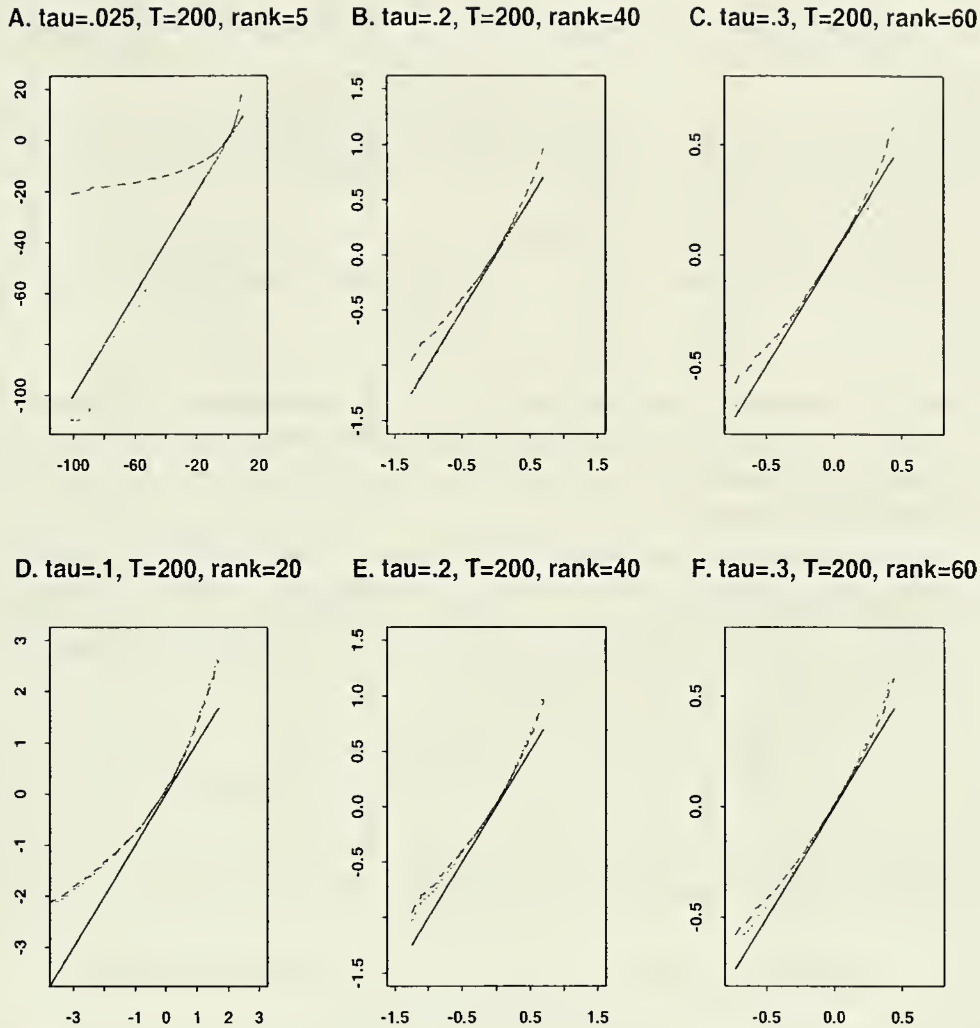

Figure 2: **Displays A-C: QQ-plot of Extreme and Central Rank Approximations.** The dashed line "- - -" is the central approximation, and the dotted line "......" is the extreme rank approximation. The true quantiles of the exact sampling distribution are depicted by the solid line "——". The central rank approximation varies from very bad to bad for low quantiles $\tau = .025$ and $\tau = .2$ and becomes comparable to the extreme rank approximation only at $\tau = .3$. **Displays D-F: QQ-plot of Intermediate and Central Rank Approximations.** The dotted line "......" now denotes the intermediate rank approximation. The theoretical central and intermediate rank approximations have approximately the same performance for $\tau = .1, .2, .3$, (using $m = 2, 1.5, 1.25$). The practical advantage of the intermediate rank approximation is the parsimony and ease of estimating nuisance parameters. [Replications $= 10,000$. QQ plots are over the 99% range.]

For fixed k, the limit laws of type 1- 3, were identified in the literature as:

$$J_k = \begin{cases} \ln \Gamma_k, & \text{for type 1 tails,} \\ -\Gamma_k^{-\frac{1}{\alpha}}, & \text{for type 2 tails,} \\ \Gamma_k^{\frac{1}{\alpha}}, & \text{for type 3 tails,} \end{cases} \tag{3.6}$$

with the *canonical scalings* given by:

$$\begin{array}{llll} \text{type 1:} & a_T \sim & 1/\ell[F_u^{-1}(\frac{1}{T})], & b_T = F_u^{-1}(\frac{1}{T}), \\ \text{type 2:} & a_T \sim & -1/F_u^{-1}(\frac{1}{T}), & b_T = 0, \\ \text{type 3:} & a_T \sim & 1/F_u^{-1}(\frac{1}{T}), & b_T = 0. \end{array} \tag{3.7}$$

Note that when $k = 1$, the type 1 law in (3.6) is called Gumbell, type 2- Frechet, type 3 - Weibull. Typically, the results state the distribution functions of J_k, but more recent treatments formulate the results in the above form (e.g. Example 4.2.5 in [26]), which helps explain our results.

Intermediate Rank Statistics. One of most general and fairly recent treatments of the intermediate order statistics is the work of Dekkers and de Haan[21]. Using slightly stronger restrictions on the tails, discussed below, they found that the limit laws are normal, but the limiting variance *depends on the extremal tail types* through the tail index ξ, as $k = [\tau T] \to \infty$ (Theorem 3.1):

$$\left[\frac{\sqrt{\tau T}}{F_u^{-1}(2\tau) - F_u^{-1}(\tau)}\right] \left(U_{([\tau T])} - F_u^{-1}(\tau)\right) \xrightarrow{d} N\left[0, \ \frac{\xi^2}{(2^{-\xi}-1)^2}\right]. \tag{3.8}$$

The scaling a_T can conveniently be replaced by $\sqrt{\tau T}/(U_{(2[\tau T])} - U_{([\tau T])})$ without affecting the result, and operationalizing the inference.

4 The Extremal Regression Quantile Models

Here we construct the linear models of low (extremal) conditional quantiles, which allow flexible covariate effects on the distribution, and coherently combine the tail conditions leading to (i) non-degenerate asymptotic distributions *and* congenial inference procedures, (ii) good approximations to the sampling distributions, and (iii) a framework suitable for inference about tails in shape-heteroscedastic models.

4.1 Model 1: Tail Homogeneity

Consider a probability space $(\Omega, \mathcal{F}, P)$, possibly indexed by T. To impose a constructive tail condition, define a *reference error* term as

$$U \equiv Y - X'\beta_r, \tag{4.9}$$

where $x \mapsto x'\beta_r$ is the *reference line* chosen so that the error U satisfies the tail homogeneity condition (i) in assumption 1. The existence of such a line is an assumption; the examples below highlight its constructive role.

In the *bounded support case*, it is convenient to choose the reference line as

$$\beta_r \equiv \beta(0), \tag{4.10}$$

so that $U \equiv Y - X'\beta(0) \geq 0$ has the end-point 0 by construction. (In the unbounded support case, $x'\beta(0) = -\infty$ and is not suitable as a reference line).

Assumption 1 (Model 1: Tail Homogeneity) *In addition to linearity (3.2):*

(i) *there is a real-valued U and reference line $x'\beta_r$ of the form (4.9)-(4.10) s.t.*

$$F_U(z|x) \sim F_u(z)$$

as $z \searrow -\infty$ (infinite support) or as $z \searrow 0$ (finite support), uniformly in $x \in \mathbf{X}$. F_u is a distribution function with type 1,2, or 3 tails.

(ii) *the support of X is (or trimmed to) a compact subset $\mathbf{X}$ of $\mathbb{R}^d$.*

(iii) *the distribution function of X_t, F_X, with support $\mathbf{X}$ and mean μ_X, is nondegenerate in $\mathbb{R}^d$. The first component of X is 1. $\mu_X = (1, 0, ...)$ w.l.o.g.*

Assumption 1-(ii), compactness, is essential; otherwise, the limits may change depending on the tail behavior of X. 1-(iii) precludes non-degeneracies.

Assumption 1-(i) requires the tails of the suitably defined error term U to be in the domain of the minimum attraction which is fairly broad with rare exceptions (section 3.5.) In this sense the *model* is *distribution-free.* 1-(i) also requires the tail of the conditional distribution function of U to be approximately independent of X. This incorporates the case of independent U and X as *strictly* a special case. Indeed, 1-(i) requires only that there is a reference error U in (4.9) such that the extremal (small) values of U are approximately independent of X. This allows general global dependence of U on X, such as shape heteroscedasticity.

Example 4.1 (Classical linear model) Suppose the quantile function is

$$Q_Y(\tau|X) = X'\alpha + F_U^{-1}(\tau), \tag{4.11}$$

which corresponds to the model $Y = X'\alpha + U$, where U is independent of X and e.g. $EU = 0$. This clearly is a *special case* of Model 1 with the reference line $x'\alpha$. Yet this example is narrow and "trivial" in the sense that the extremal features are determined by the central features of the distribution, and there is "nothing to estimate" (all slope coefficients $\beta_{-1}(\tau)$ equal α_{-1}.) To defend the "trivial" model, note it underlies much of the (central) quantile regression inference, Koenker and Bassett[48], because it often *plausibly* approximates the exact distribution of regression quantiles, even though the model itself is unrealistic (Koenker and Hallock[50]).

Example 4.2 Consider the bounded support case. The 0-th quantile function is $Q_Y(0|X) = X'\beta(0)$, which is our reference line $X'\beta_r$. By assumption 1(i),

$$P(Y - X'\beta(0) \leq l|X) \sim F_u(l) \equiv \tau_l, \text{ as } l \searrow 0,$$

which implies that the paths of the extremal quantile functions $x \mapsto x'\beta(\tau_l)$ are *approximately* parallel to that of $x \mapsto x'\beta(0)$. This model is not "trivial" in the sense of Example 4.1, because the extremal quantiles and the reference line are determined only by the extremal features of the conditional distribution. The model does not restrict other quantiles, allowing for general shape- heteroscedasticity. For example, a collection of quantile curves $\{x \mapsto Q_Y(q|x), q \in \mathcal{C}\}$ with the central indices $\mathcal{C}$ may have complicated non-parallel paths, allowing for complicated effects of covariates on the conditional density shape (kurtosis, skewness, and other effects), as in Figure 3. Consequently, this model *does not admit reductions* to non-regression models by removing a conditional location and scale from Y.

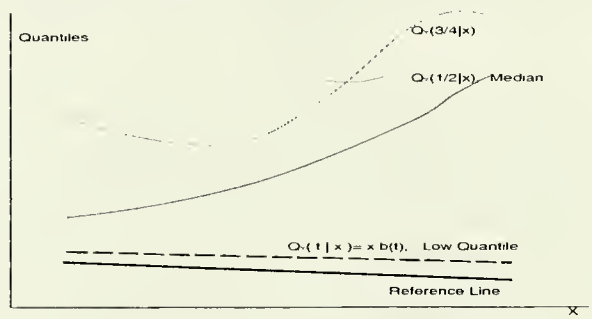

Figure 3: Example 4.2: Extremal conditional quantile function $x \mapsto x'\beta(\tau)$ is approximately parallel to the reference line $x \mapsto x'\beta_r$ (equal to the minimal quantile $x'\beta(0)$ in the bounded support case). Other quantile functions are unrestricted, allowing for complicated forms of global heteroscedasticity. The model does not admit the reduction to a non-regression model by removing the conditional median (or mean) and/or scale from Y variable.

Example 4.3 Consider the unbounded support. For some reference line $x \mapsto x'\beta_r$, by assumption 1(i),

$$P(Y - X'\beta_r \leq l|X) \sim F_u(l) \equiv \tau_l, \text{ as } l \searrow -\infty,$$

which implies that the paths of the extremal quantile functions $x \mapsto x'\beta(\tau_l)$ are approximately parallel to that of $x \mapsto x'\beta_r$. This model is also not "trivial" in the sense of example 4.1, because the extremal quantiles are determined only by the extremal features of the conditional distribution. As in example 4.2, the model does not restrict any other features of the distribution, allowing for general forms of global heteroscedasticity. Thus it is irreducible to a non-regression model.

Note that examples 4.2 and 4.3 demonstrate that the linear location-scale models are neither implied by Model 1 nor imply Model 1. Thus Model 1 is of its own nature, crafted to yield non-degenerate, parsimonious limits. Unlike the location-scale models, Model 1 admits general global heteroscedasticity, allowing covariates to affect the shape of the conditional distribution.

4.2 Model 2: Congenial Tail Heterogeneity

We suggest a model that, while flexibly accounting for the dependence of the tail on covariates, exhibits simplicity, enabling an explicit, practical limit theory for both the extreme and intermediate rank sample regression quantiles.

Assumption 2 (Model 2: Congenial Tail Hetrogeneity) Suppose assumption 1 holds, except 1-(i) is replaced by the following tail condition:

$$F_U(z|x) \sim K(x) \cdot F_u(z), \ \text{as } z \searrow -\infty \text{ or } z \searrow 0, \tag{4.12}$$

uniformly in $x \in \mathbf{X}$, F_u has type 1-3 tails, $K(\cdot)$ is assumed to be a positive continuous function on $\mathbf{X}$, bounded above and away from zero, normalized so that $K(\mu_X) = 1$ (or at any other reference point $x^r \in \mathbf{X}$).

Just like Model 1, Model 2 is *distribution-free*, since F_u is not assumed to be parametric, and it allows the general (shape) forms of global heteroscedasticity. Unlike Model 1, Model 2 allows for *richer* effects of covariates on tails.

The imposed tail condition may seem an unconventional way to introduce heteroscedasticity. Yet, in many regards, it is more flexible and constructive than the conventional location-scale modeling, as explained below. The proposed modeling strategy is motivated by the closure of the domains of minimum attraction under tail equivalence, and is fully *consistent with* linearity.

Indeed, Lemma 10, characterizes this model in detail: **(i)** implications for the quantile coefficients of the linear model, **(ii)** limits of ratios of spacings between the conditional quantile functions, and **(iii)** many other properties needed for inference. Importantly, we deduced that the linearity assumption and (4.12) *jointly* imply that $K(\cdot)$ can be represented as

$$K(x) = \begin{cases} e^{-x'\mathbf{c}} & \text{for type 1 tails,} \\ (x'\mathbf{c})^{\alpha} & \text{for type 2 tails,} \\ (x'\mathbf{c})^{-\alpha} & \text{for type 3 tails,} \end{cases} \tag{4.13}$$

where $\mu_X'\mathbf{c} = 1$ for type 2 and 3 tails, and $\mu_X'\mathbf{c} = 0$ for type 1 tails. In Model 1, $\mathbf{c} = \mathbf{0}$ for type 1 tails, and $\mathbf{c} = (1, 0, ...)' = \mathbf{e}_1'$ for type 2 and 3 tails. We call $\mathbf{c}$ the *tail heterogeneity index*. It measures the *strength* with which X *shift the tails* of error terms U. Note that $x'\mathbf{c} > 0$ uniformly on $\mathbf{X}$ for types 2 and 3 by assumption.

It is plausible that (potentially) the non-parametric function $K(\cdot)$ in (4.12) is in fact a transformation of the linear index $x'\mathbf{c}$ determined by the tail index ξ. Recall $\xi = 0$ (for type 1 tails) and $\xi = 1/\alpha$ and $-1/\alpha$ for type 2 and 3 tails, respectively. This assumption leads to parsimonious, convenient limits for regression quantiles.

The following examples illustrate the model's flexibility.

Example 4.4 (Linear Location-Scale Model) Assume for $X'\gamma > 0$ a.s.

$$Q_Y(\tau|x) = x'\alpha + x'\gamma \cdot F_V^{-1}(\tau), \tag{4.14}$$

corresponding to the location model $Y = X'\alpha + X'\gamma \cdot V$, where V is independent of X and, say, has mean 0 and variance 1. Assume F_V has the extremal tail type with $\xi \neq 0$. Then for the reference line $x'\alpha$ and $U \equiv Y - X'\alpha = X'\gamma \cdot V$

$$P(X'\gamma \cdot V \leq l|X) \sim (X'\gamma)^{-1/\xi} \cdot F_V(l), \ \text{as } l \searrow -\infty,$$

so the conditions of Model 1 are satisfied with $F_u \equiv F_V$. The location-scale model imposes two stringent restrictions: (i) the extremal features of the distribution are largely determined by the (central) location and scale parameters: $\beta(\tau) = \alpha + \gamma \cdot F_V^{-1}(\tau)$, and (ii) the covariates are limited to affect only the location and scale of the conditional distribution, precluding the shape effects like skewness or kurtosis.

Example 4.5 Model 2 requires that for some reference line $x'\beta_r$

$$P(Y - X'\beta_r \leq l|X) \sim K(x) \cdot F_u(l), \text{ as } l \searrow 0 \text{ or } -\infty,$$

which implies that the paths of the extremal quantile functions $x \mapsto x'\beta(\tau_l)$ are no longer parallel to that of $x \mapsto x'\beta_r$. (The crossing of lines is precluded because the assumption is consistent with linearity, Lemma 10). This model is not as restrictive as example 4.4. First, the extremal quantiles (and the reference line) are determined only by the extremal features of the conditional distribution. Second, the model allows for general global heteroscedasticity – the *entire* shape of the conditional density may change with covariates (scale, skewness, etc), including the tails.

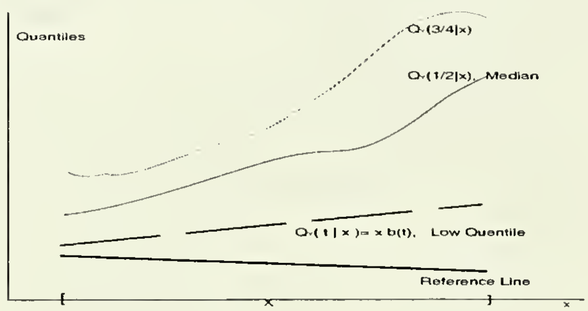

Figure 4: Example 4.5. Extremal quantile functions $x \mapsto x'\beta(\tau)$ are no longer approximately parallel to the reference line $x \mapsto x'\beta_r$, over **X**, allowing the tail heteroscedasticity. Other quantile functions are unrestricted, allowing for complicated forms of global heteroscedasticity as well. The extremal features of the model, including the reference lines, are not determined by the central features.

This discussion concludes the construction of our models. In principle, it should be possible to further relax the modeling assumptions, particularly in the nonparametric direction, but a good deal of *caution* is needed to assure the *joint coherency* of the tail conditions, functional dependence of the quantile curves on regressors, and non-degeneracy of the limit distributions. Note that the obtained models are coherent, flexible, distribution-free, lead to non-degenerate, parsimonious limit distributions, and provide a convenient framework for inference about the tails.

5 Asymptotics under Extreme Ranks

Recall that the approximation concept of extreme ranks requires $\tau T \to k > 0$. Here we state the distribution results for Models 1 and 2 and explain their barest essence, while leaving proofs and some generalizations to the appendix.

5.1. A Sketch. First, obtain a finite-dimensional (*fidi*) weak limit $Q_\infty(\cdot)$ of the finite-sample, suitably scaled, objective functions $\{Q_T(\cdot)\}$. Q_∞ is defined by a point process that "counts" the "extremal events." Then, the normalized regression quantile statistic, $\hat{Z}_T$, an argmin of Q_T, will converge in distribution to a random variable Z_∞, the argmin of Q_∞, by convexity of $\{Q_T\}$ and Q_∞.

For brevity, we confine our discussion to type 3 tails. Consider the statistic

$$\hat{Z}_T \equiv a_T\big(\hat{\beta}(\tau) - \beta(0)\big),$$

where a_T is the *canonical* scaling in section 3.6, defined in terms of the function F_u in assumption 1 or 2. $\hat{Z}_T$ optimizes the rescaled by a_T objective function in (3.4):

$$Q_T(z) \equiv \sum_{t=1}^{T} a_T\rho_\tau\big(U_t - X_t'z/a_T\big) - a_T(\tau U_t), \tag{5.15}$$

where $U_t \equiv Y_t - X_t'\beta(0) \geq 0$ and $z = a_T(\beta - \beta_\tau)$. We subtracted the "smoother" $\sum_t \tau U_t a_T$, which brings a key continuity property and stabilizes Q_T. Clearly this does not affect the argmin $\hat{Z}_T$. [The "smoother" for type 1-2 tails is more involved; Lemma 1. Incidentally, the conventional central rank stabilization by $\sum_t \rho_\tau(U_t a_T)$ is bad, for it sends the objective to $+\infty$, let alone continuity.] Hence

$$Q_T(z) = -T\tau\bar{X}'z - \sum_{t=1}^{T} 1\big(U_t a_T \leq X_t'z\big)\big(U_t a_T - X_t'z\big). \tag{5.16}$$

This function is convex. Notably, it is *constructed* as a *continuous functional* of the point process defined next. The fi-di distribution of Q_T is defined by that of $\{Q_T(z_j), j \leq l\}$ for any finite $(z_j, j \leq l)$. Since $\bar{X} \xrightarrow{p} \mu_X$, and $\tau T \to k$ (for $j \leq l$) :

$$Q_T(z_j) = -k\mu_X' z_j - \sum_{t=1}^{T} 1(U_t a_T \leq X_t z_j)\big(U_t a_T - X_t' z_j\big) + o_p(1). \tag{5.17}$$

The limit behavior of Q_T is *determined* by the *point process* $\widehat{\mathbf{N}}$ that assigns mass to measurable sets A by:

$$\widehat{\mathbf{N}}(A) = \sum_{t=1}^{T} 1(\{a_T U_t, X_t\} \in A), \text{ for } A \subset E \equiv [0,\infty) \times \mathbf{X}.$$

The point process $\widehat{\mathbf{N}}$ is a measure defined by its random points (atoms) $(a_T U_t, X_t, t \leq T)$ (See Definition A.1, B.1). We find that $Q_T(\cdot)$ is an *integral* of a residual function with respect to the point process, which seems to be special to this problem.

Point process theory is the bread-and-butter of extreme value theory,[8] [26], and is useful here. Indeed, a Lebesgue-Stieltjes integral $\int g dN$ of N with points $\{X_i\}$ is :

$$\int g(x)dN(x) \equiv \sum_i g(X_i).$$

[8] Point process theory was developed by Kallenberg[44], Resnick[65] and others in considerable generality. Applications are numerous in statistics. For example, Feigin and Resnick[29] approximate the constraints of the linear programming estimators; also Knight[46] ; Emrechts et al.[26] and Resnick[65] show how point processes may be used in related applications, particularly the exceedance processes, extremal processes, and record values.

Convergence of such integrals, for continuous maps $x \mapsto g(x)$ that vanish outside compact sets, metrizes weak convergence of point processes (Definition A.2.) So we represent $(Q_T(z_j), j \leq l)$ as an integral

$$Q_T(z_j) = -k\mu_X' z_j + \int (l - x'z_j)^- d\widehat{\mathbf{N}}(l, x) + o_p(1), \tag{5.18}$$

where $(l - x'z)^-$ is the "residual" function. Lemma 2 proves $\big(Q_T(z), j \leq l\big)$ is a *continuous map* of the *point process* $\widehat{\mathbf{N}}$, so its weak limit is determined by that of $\widehat{\mathbf{N}}$. Notably, to obtain continuity for various tail types, the construction of point process $\widehat{\mathbf{N}}$ requires a careful choice of the underlying topological space E (and additional transformations of Q_T for types 1 and 2).

The weak limit (Def. A.2) of $\widehat{\mathbf{N}}$ in Model 2 is a Poisson process $\mathbf{N}$, Lemma 6:

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

where $\{J_i, \mathcal{X}_i\}$ are random points defined as

$$\begin{aligned} &\Big(J_i,\ \mathcal{X}_i,\ \ i \geq 1\Big) \stackrel{d}{=} \Big(\mathcal{X}_i'\mathbf{c} \cdot \Gamma_i^{\frac{1}{\alpha}},\ \mathcal{X}_i,\ \ i \geq 1\Big), \\ &\Gamma_i \equiv \mathcal{E}_1 + ... + \mathcal{E}_i,\ \ i \geq 1, \end{aligned} \tag{5.19}$$

where $\{\mathcal{E}_i\}$ are i.i.d. exponential random variables with mean 1, $\{\mathcal{X}_i\}$ are i.i.d. with law F_X, distributed independently of $\{\mathcal{E}_i\}$, and $\mathbf{c}$ is the tail heterogeneity parameter. In Model 1, because $\mathbf{c} = (1, 0, ...)$, a natural simplification occurs:

$$\mathcal{X}_i'\mathbf{c} = 1, \forall i. \tag{5.20}$$

The first result, explained in Lemma 6, is not self-evident, while (5.20) is fairly intuitive. Note that $\widehat{\mathbf{N}}(A) = \sum_{i \leq T} 1(\{a_T U_{(i)}, X_{\{i\}}\} \in A)$, where $U_{(i)}$ is i-th rank error, and $X_{\{i\}}$ is the corresponding covariate. Vector $(a_T U_{(i)}, i \leq q) \xrightarrow{d} (\Gamma_i^{1/\alpha}, i \leq q)$ (Section 3.6), and is asymptotically independent of X_i by Assumption 1-(i) in Model 1, which explains the form of (5.19) and (5.20) for Model 1. (This is not a proof). Lemmas 5-6 provide the proof for Model 2 (and 1 by implication) using the Kalenberg's theorem, Meyer's conditions, and a series of compositions and transformations of a canonical Poisson process (Def. A.4 provides a background).

We conclude that the fidi weak limit of $\{Q_T\}$ is

$$Q_\infty(z) \equiv -k\mu_X' z + \int (j - x'z)^- d\mathbf{N}(j, x) \equiv -k\mu_X' z + \sum_{i=1}^{\infty} (J_i - X_i'z)^-.$$

Therefore, we obtain by convexity Lemma 1 in the appendix (Theorem 5 in Knight[46]) the limit distribution for $\hat{Z}_T$, provided $Q_\infty(\cdot)$ has a unique argmin a.s. and is finite on an open non-empty set (verified in Lemma 2 and 11). Hence

$$a_T\Big(\hat{\beta}(\tau) - \beta(0)\Big) \xrightarrow{d} Z_\infty \equiv \underset{z}{\text{argmin}}\ Q_\infty(z). \tag{5.21}$$

Finally, Lemma 10 shows that $a_T(\beta(0)-\beta(\tau)) \to k^{\frac{1}{\alpha}}\mathbf{c}$ in Model 2 ($\mathbf{c}=\mathbf{e}_1=(1,0,...)'$ in Model 1) so that

$$a_T\Big(\hat\beta(\tau)-\beta(\tau)\Big) \xrightarrow{d} -k^{\frac{1}{\alpha}}\mathbf{c}+Z_\infty.$$

5.2. Results for Models 1 and 2. The above discussion hopefully provided an intuitive explanation of the foregoing formal results for Models 1 and 2 (Theorems 1 and 2). The proofs are in the appendix. To state the result, suppose we have l sequences $\{\tau_i, i\le l\}$ such that $\tau_i T\to k_i$, so we index the normalized regression quantile statistic as $\hat Z_T(k_i)$, for both $T<\infty$ and $T=\infty$. Define

$$\begin{aligned} \hat Z_T(k) &\equiv a_T\left(\hat\beta(\tau)-\beta_r-b_T\mathbf{e}_1\right), && \text{for type 1 tails,}\\ \hat Z_T(k) &\equiv a_T\left(\hat\beta(\tau)-\beta_r\right), && \text{for type 2 \& 3 tails.}\end{aligned}$$

Also define the centered statistic

$$\hat Z_T^c(k)\equiv a_T\Big(\hat\beta(\tau)-\beta(\tau)\Big).$$

The canonical constants (a_T,b_T) are defined in (3.7) in terms of functions F_u, which are defined in Assumptions 1 and 2 along with the error term U_t and β_r.

The key point process, $\widehat{\mathbf{N}}(\cdot)=\sum_{t\le T}1(\{a_T(U_t-b_T),X_t\}\in\cdot)$ weakly converges (Def. A.2) to $\mathbf{N}(\cdot)=\sum_{i\ge 1}1(\{J_i,\mathcal{X}_i\}\in\cdot)$ by Lemma 4-6, with points $\{J_i,\mathcal{X}_i\}$ defined as:

$$\Big(J_i,\ \mathcal{X}_i,\ i\ge 1\Big)\stackrel{d}{=}\begin{cases}(\ln(\Gamma_i)+\mathcal{X}_i'\mathbf{c}, & \mathcal{X}_i) \quad \text{for type 1,}\\ (\Gamma_i^{-1/\alpha}\mathcal{X}_i'\mathbf{c}, & \mathcal{X}_i) \quad \text{for type 2,}\\ (\Gamma_i^{1/\alpha}\mathcal{X}_i'\mathbf{c}, & \mathcal{X}_i) \quad \text{for type 3,}\end{cases}\qquad i\ge 1 \tag{5.22}$$

where $\{\Gamma_i, i\ge 1\}=\{\sum_{j\le i}\mathcal{E}_j, i\ge 1\}$; $\{\mathcal{E}_j\}$ is an i.i.d. sequence of unit-exponential variables; $\{X_i\}$ is an i.i.d sequence with law F_X. In Model 1, the dependence between J_i and $\mathcal{X}_i$ naturally disappears in view of assumption 1-(i):

$$\begin{aligned}\mathcal{X}_i'\mathbf{c}&=0 \text{ for type 1 tails}, \forall i,\\ \mathcal{X}_i'\mathbf{c}&=1 \text{ for type 2 \& 3 tails}, \forall i.\end{aligned} \tag{5.23}$$

Theorem 1 (Extreme Rank Asymptotics in Model 1) *Suppose Assumption 1 and that (a) $\{Y_t,X_t\}$ is an i.i.d. or stationary sequence, satisfying the Meyer conditions, Lemma 6; (b) at least one component of X is absolutely continuous, if $d\ge 2$. Then as $\tau T\to k, T\to\infty$, $(k=k_1,...,k_l)$, for a.e. $k>0$*

$$\hat Z_T(k)\xrightarrow{d} Z_\infty(k)\equiv \underset{z\in\mathbb{R}^d}{\operatorname{arginf}}\Big[-k\mu_X' z+\int l(u,x'z)d\mathbf{N}(u,x)\Big],$$

where $l(u,v)\equiv 1(u\le v)(v-u)$, and the distribution of points $\{J_i,X_i\}$ of $\mathbf{N}$ is defined in (5.22)-(5.23). Furthermore, $\big(\hat Z_T(k_i), i\le l\big)\xrightarrow{d}\big(Z_\infty(k_i), i\le l\big)$,

$$\hat Z_T^c(k)\xrightarrow{d} Z_\infty^c(k)\equiv Z_\infty(k)-\mathbf{c}(k),$$

and $\big(\hat Z_T^c(k_i), i\le l\big)\xrightarrow{d}\big(Z_\infty^c(k_i), i\le l\big)$, where $\mathbf{c}(k)=\ln k\ \mathbf{e}_1$, for type 1, $-k^{\frac{-1}{\alpha}}\mathbf{e}_1$, for type 2, and $k^{\frac{1}{\alpha}}\mathbf{e}_1$, for type 3 tails.

The asymptotic distribution is that of the random variable $Z_\infty(k)$. The density of Z_∞ can be simulated. Analytical formulae for this density are given in Remark 5.1. Remark 5.2 explores the connections to the classical results.

Theorem 2 (Extreme Rank Asymptotic in Model 2) *Suppose the assumptions of Theorem 1, with Assumption 2 replacing Assumption 1. The statement of Theorem 1 remains valid, with points $\{J_i, \mathcal{X}_i\}$ of $\mathbf{N}$ defined in (5.22) and the centering constants $\mathbf{c}(k)$ defined as follows: $\mathbf{c}(k) = \ln k\mathbf{e}_1 + \mathbf{c}$, for type 1, $-k^{\frac{-1}{\alpha}}\mathbf{c}$, for type 2, $k^{\frac{1}{\alpha}}\mathbf{c}$, for type 3 tails.*

We stated the result for Model 2 separately in order to emphasize the model's congeniality. It exhibits simplicity, while allowing flexible dependence of the tail on covariates. Both Models 1 and 2 allow us to fully characterize the limit process $\mathbf{N}$, which defines a parsimonious limit distribution in terms of only two parameters – the tail index ξ and tail heterogeneity index $\mathbf{c}$ (known for Model 1). The scaling constants a_T are of the same form for both models.

Note that Theorems 1-2 allow for weakly dependent data as well. The point of this paper (by far) is not about dependent data, but since the proof takes only an additional half-page, we thought it shameless not to state the result. The imposed Meyer conditions require strong-mixing and no-clustering of the data sequence. See Lemma 5. Notably, because rare events separate in time, all the limits are *identical* to those of an independent sequence. This is analogous to the results of Robinson[66] on kernel estimation, where the relevant local events are asymptotically independent.

Again, it is *not* our goal to *dwell on technicalities*, but it is reasonable to examine the density of Z_∞. If it is simple (it's not), it should be very useful in practice.

Remark 5.1 (Asymptotic Density) Let $\mathcal{H}$ be the set of all d-element permutations of integers $1, 2, \ldots$. 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. $\{J_t\}$ are absolutely continuous, conditional on $\{\mathcal{X}_t\}$. Conclude, mimicking computations of the gradient and finite-sample density for quantiles in Koenker and Bassett[48]: a. An argmin of Q_∞ takes the form $z = \mathcal{X}(h)^{-1}J(h)$ (passage through d-points) and it is unique iff

$$\zeta_h(z) \equiv (k\mu_X - \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}_t'z)\mathcal{X}_t)'\mathcal{X}(h)^{-1} \in \mathcal{D} \equiv (0,1)^d, \tag{5.24}$$

and is non-unique if $\zeta_h(z) \in \partial\mathcal{D}$. If $\mathcal{X}_t$ has an absolutely continuous component, $\det \mathcal{X}(h)$ is absolutely continuous ($\det \mathcal{X}(h)$ is a volume of the parallelogram formed by $\mathcal{X}(h)$) so that $\zeta_h(z) \in \partial\mathcal{D}$ w.p 0; b. Given (5.24), the density of Z_∞ is

$$f_{Z_\infty}(z) = E\Big[\sum_{h\in\mathcal{H}} f_{J(h)|\mathcal{X}(h)}(\mathcal{X}(h)'z) \cdot |\det \mathcal{X}(h)| \cdot \; P\left(\zeta_h(z) \in \mathcal{D} | \{\mathcal{X}_t\}, h\right)\Big],$$

where $f_{J(h)|\mathcal{X}(h)}$ is the conditional on $\mathcal{X}(h)$ joint density of $J(h)$.

Remark 5.2 (Relating to Classical Theory) $P(\zeta_h(z) \in \mathcal{D}|\{\mathcal{X}_t\}, h)$ is hard to obtain explicitly. It simplifies in the no-X case, $X = 1$, $P\left(\zeta_h(z) \in \mathcal{D}|h\right) = 1$, if $h = \lceil k \rceil$, 0, if not. k must be a non-integer for uniqueness. Hence $f_{Z_\infty}(z) = f_{J_{(\lceil k \rceil)}}(z)$, which is the limiting distribution of the $\lceil k \rceil$-th order statistics in the i.i.d. samples.

6 Asymptotics under Intermediate Ranks

The intermediate rank concept requires $\tau \to 0, \tau T \to \infty$. In this section we first state the results for Models 1 and 2, followed by a brief explanation.

6.1 Results for Models 1 and 2. In addition to assumptions 1 and 2, we require existence of density $f_U(\cdot|x)$ or, equivalently, of $\partial F_U^{-1}(\tau|x)/\partial\tau \equiv x'\partial\beta(\tau)/\partial\tau$. This density should posses enough smoothness. We also need the conditional density tail-equivalence, an assumption that strengthens the tail equivalence of the conditional distribution functions in Models 1 and 2.

Assumption 3 (Density Conditions) *(i) In the Model 1, as* $\tau \searrow 0$

$$\frac{\partial F_U^{-1}(\tau|x)}{\partial\tau} \sim \frac{\partial F_u^{-1}(\tau)}{\partial\tau}, \tag{6.25}$$

uniformly in $x \in \mathbf{X}$, *and in Model 2,*

$$\frac{\partial F_U^{-1}(\tau|x)}{\partial\tau} \sim \frac{\partial F_u^{-1}(\tau/K(x))}{\partial\tau} \tag{6.26}$$

(ii) $\partial F_u^{-1}(\tau)/\partial\tau$ *is regularly varying at 0 with exponent* $-\xi-1$, *cf. section 3.5 (denote* $\partial F_u^{-1}(\tau)/\partial\tau \in \mathcal{R}_{-\xi-1}$*).*

Assumptions 3(i) and 3(ii) are both constructive and general. Assumption 3(i) is a stronger density analog of the tail equivalence conditions imposed on $F_U(\tau|x)$ in Assumptions 1 and 2. Assumption 3(ii) is an analytical smoothness condition on the density. It was first proposed in the non-regression context by Dekkers and de Haan[21], who also show that the exceptions among the smooth distributions are rare.[9]

Fix a reference index sequence $\{\tau\}$ such that $\tau \searrow 0$ and $\tau T \to \infty$. Consider l sequences $\{\tau l_i\}, i \le k$, and define $\mathbf{Z}_T \equiv \left(a_T(l_i)[\hat\beta(l_i\tau) - \beta(l_i\tau)], i \le k\right)$, where

$$a_T(l) \equiv \sqrt{\tau l T}/\mu_X' \left(\beta(ml\tau) - \beta(l\tau)\right),$$

for positive l and $m > 0, \neq 1$. Set $a_T \equiv a_T(l)|_{l=1}$.

Theorem 3 (Intermediate Rank Asymptotics in Model 1) *Suppose Assumptions 1 and 3 hold, and that* $\{Y_t, X_t\}$ *is an i.i.d. or stationary series, satisfying the conditions of Lemma 9, then as* $\tau T \to \infty, \tau \searrow 0$

$$a_T\left(\hat\beta(\tau) - \beta(\tau)\right) \xrightarrow{d} N(0, \mathbf{V}), \quad \mathbf{V} \equiv Q_X^{-1}\frac{\xi^2}{(m^{-\xi}-1)^2}, \tag{6.27}$$

where $Q_X \equiv EXX'$. $\mathbf{Z}_T \xrightarrow{d} N(0,\Omega)$, $\Omega_{ij} \equiv \mathbf{V} \times \min(l_i,l_j)/\sqrt{l_il_j}$. *Furthermore,* $a_T(l)$ *can be replaced by* $\sqrt{\tau l T}/\bar X'\left(\hat\beta(ml\tau) - \hat\beta(l\tau)\right)$ *without affecting the result.*

Remark 6.1 It may be useful to have the same normalization a_T in place of $a_T(l)$ for the joint convergence in distribution. This is possible by noting that $a_T/a_T(l) \to l^{-\xi}/\sqrt{l}$. Then $\left(a_T(\hat\beta(l_i\tau) - \beta(l_i\tau)), i \le n\right) \xrightarrow{d} N(0,\Sigma)$, $\Sigma_{ij} \equiv \Omega_{ij}(l_il_j)^{-\xi}/\sqrt{l_il_j}$.

[9] To see the plausibility, take near-algebraic and differentiable near finite or infinite lower end-points distributions: $F_u(z) = Cz^{-1/\xi}(\ln z)^K$ as $z \searrow 0$ or $F_u(z) = C(-z)^{-1/\xi}(\ln[-z])^K$ as $z \searrow -\infty$, $K \in \mathbb{R}$.

Theorem 4 (Intermediate Rank Asymptotics in Model 2) Suppose Assumptions 2 and 3 (where appropriate) hold, and that $\{Y_t, X_t\}$ is an i.i.d. or stationary series, satisfying conditions of Lemma 9. The results of Theorem 3 remain valid, with the variance matrix $\mathbf{V}$ *taking the form*

$$\mathbf{V} \equiv \mathcal{Q}_H^{-1} \mathcal{Q}_X \mathcal{Q}_H^{-1} \frac{\xi^2}{(m^{-\xi} - 1)^2},$$

$\mathcal{Q}_H \equiv E[H(X)]^{-1} XX'$, *where* $H(x) \equiv x'\mathbf{c}$ *for type 2 and 3, and 1 for type 1 tails.*

Because the intermediate rank theory exploits the *extremality* of the relevant events, the limit is defined only by the tail parameters. *Unlike* the extreme rank approximation, the condition relies on the relative abundance of the relevant tail events, which leads to normality. This produces a convenient theory, on which an effective and practical inference can be based, as further discussed in Section 7. Theorems 3 and 4 create a basis for a series of results that give consistent estimates of the important tail parameters (Section 7, Remark 6.2).

Lastly, dependent data is handled as well. Even though this is not the main focus of the paper, the proof is very short using the local CLT of Robinson[66](see Lemma 9). Notably, the resulting limit is the same as for the independent data (due to the separation of the tail events in time).

6.2. A Sketch. This provides a brief explanation of the result. The key difficult steps are treated in the appendix, Lemmas 7-10. Our approach substantively differs from the ingenious proof of Dekkers and de Haan[21] for the unconditional case. They use the Renyi representation of order statistics, an approach that can not be applied here.

The normalized statistic $\hat{Z}_T \equiv a_T\big(\hat{\beta}(\tau) - \beta(\tau)\big)$ minimizes

$$Q_T(z) \equiv \frac{a_T}{\sqrt{\tau T}} \Big(\sum_{t=1}^{T} \rho_\tau\big(Y_t - X_t'\beta(\tau) - \frac{X_t z}{a_T}\big) - \rho_\tau\big(Y_t - X_t'\beta(\tau)\big)\Big).$$

We seek to find its finite-dimensional weak limit $Q_\infty(\cdot)$, so that by convexity we may conclude $\operatorname{argmin} Q_T(z) \xrightarrow{d} \operatorname{argmin} Q_\infty(z)$. Write

$$Q_T(z) \equiv \left[\frac{-1}{\sqrt{\tau T}} \sum_{t=1}^{T} (\tau - 1[Y_t \le X_t\beta(\tau)])X_t'z\right] + \left[\frac{1}{\sqrt{\tau T}} \sum_{t=1}^{T} \mu_t(z)[[Y_t - X_t\beta(\tau)]a_T - X_t'z]\right] \equiv W_T' \, z \; + \; G_T(z),$$

where $\mu_t(z) \equiv \big(1(Y_t \le X_t'\beta(\tau)) - 1(Y_t \le X_t'\beta(\tau) + X_t'z/a_T)\big)$.

Lemma 8, eq. (C.47), proves that Assumptions 3(i) and 3(ii) *delicately* imply

$$E\ \mu_t(z) = O\left(f_u(F_u^{-1}(\tau)) \cdot a_T^{-1}\right). \tag{6.28}$$

This equality is *not* self-evident; in fact, it is counter-intuitive since a_T may be converging to zero. Lemma 8, eq. (C.52), also shows that due to Assumption 3-(ii)

$$f_u(F_u^{-1}(\tau)) \cdot \left[\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau}\right] \longrightarrow \frac{m^{-\xi} - 1}{-\xi}, \tag{6.29}$$

for all $m > 0, \neq 1$, as $\tau \searrow 0$. Then, by compactness of $\mathbf{X}$,

$$\text{Var } G_T(z) = O\left(\frac{T}{\tau T} f_u(F_u^{-1}(\tau)) a_T^{-1}\right) = o(1). \tag{6.30}$$

Lemma 9 handles the dependence. Thus $G_T(z) - E\ G_T(z) \xrightarrow{p} 0$. Lemma 8, using Assumptions 3 (i) and (ii), proves that

$$E[G_T(z)] \longrightarrow \frac{1}{2} z' Q_H z \cdot \left| \frac{m^{-\xi} - 1}{-\xi} \right| \equiv \frac{1}{2} z' J(m) z, \text{ for any } z, \tag{6.31}$$

where Q_H is defined in Theorem 4. For Model 1, this simplifies $Q_H = Q_X$.

Assumption 3 is instrumental in verifying (6.31) and (6.28), which is the most difficult and important part of the proof.

By the Lindeberg CLT or Robinson's local CLT for dependent data, Lemma 9,

$$W_T \xrightarrow{d} W_\infty \equiv N(0, Q_X), \tag{6.32}$$

and $\text{Var}(W_T) \longrightarrow Q_X$. Because of the separation of the tail events in time, dependence ceases to matter in the limit. Notably, the Liapunov and other central limit theorems which require strictly more than two bounded moments *do not* apply.

Therefore, the finite-dimensional weak limit of $Q_T(\cdot)$ is

$$Q_\infty(z) \equiv W_\infty' z + \frac{1}{2} z' J(m) z. \tag{6.33}$$

Since Q_T and Q_∞ are convex and a.s. finite, and Q_∞ is uniquely minimized at $Z_\infty \equiv -J^{-1}(m)\ W_\infty = O_p(1)$, we conclude $Z_T \xrightarrow{d} Z_\infty$ by convexity Lemma 1. The joint convergence follows similarly by considering a sum of scaled objective functions, and proceeding as above.

Lastly, the scaling a_T can be replaced by its empirical analog:

$$\begin{aligned} \frac{\bar{X}'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} &\equiv \frac{\bar{X}'(\hat{\beta}(m\tau) - \beta(m\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \\ &- \frac{\bar{X}'(\hat{\beta}(\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} + \frac{\bar{X}'(\beta(m\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \xrightarrow{p} 1, \end{aligned} \tag{6.34}$$

since the first two elements on the r.h.s. are of order $O_p(\frac{1}{\sqrt{\tau T}}) = o_p(1)$.

The last display states that a population quantile spacing can be replaced by its empirical analog, which is remarkably useful for inference purposes (Section 7). This property is unexpected at a first sight, see Remark 6.2.

Remark 6.2 (Empirical Regression Quantile Spacings.) Note that (6.34) does not follow from the convergence of $\hat{\beta}(\tau)$ to $\beta(\tau)$ under intermediate ranks, because $\beta(m\tau) - \beta(\tau)$ may be converging to 0 or diverging to infinity, as $\tau \searrow 0$. Furthermore, when $\sqrt{\tau T}/\mu_X'(\beta(m\tau) - \beta(\tau)) \to 0$, $\hat{\beta}(\tau)$ does not converge to $\beta(\tau)$. Indeed, take F_u in Assumption 1 to be of type 2. Then $\mu_X'(\beta(m\tau) - \beta(\tau)) \sim \tau^{-\frac{1}{\alpha}} \mathcal{L}(\tau)$, for some slowly varying function $\mathcal{L}$ at 0. Then if $\{\tau_T\}$ satisfies $\tau = cT^{-\lambda}$, $\lambda \in (0, 1)$, and if $\frac{2\lambda}{1-\lambda} > \alpha > 0$, then $\sqrt{\tau T}/\mu_X'(\beta(m\tau) - \beta(\tau)) \to 0$. That is, divergence arises when $\tau \to 0$ fast, and the tails are sufficiently thick. Notably, (6.34) remains valid.

7 Inference

The focus of this paper is the modeling and distribution theory for the extremal (high and low) regression quantiles. It is also desirable (but not feasible) to address all practical issues related to confidence intervals, hypothesis testing, and tail inference within the present paper. These questions are partly addressed in [13], and we hope to pursue them in future work. Here, we present a brief discussion of these issues.

7.1 Quantile Spacings and Tail Inference. Estimation of the tail index is an important problem in the statistics of extreme values, as discussed in Example 2.3. The tail parameters also enter the limit distributions obtained earlier. The following results show how to estimate them by the sample *regression quantile spacings.*

Consider the following statistics

$$\varphi = \frac{x'(\hat\beta(m\tau) - \hat\beta(\tau))}{x'(\beta(m\tau) - \beta(\tau))},\ \rho_{x,\dot x,l} = \frac{x'(\beta(ml\tau) - \beta(l\tau))}{\dot x'(\beta(m\tau) - \beta(\tau))},\ \hat\rho_{x,\dot x,l} = \frac{x'(\hat\beta(ml\tau) - \hat\beta(l\tau))}{\dot x'(\hat\beta(m\tau) - \hat\beta(\tau))}.$$

Theorem 5 (Regression Quantile Spacings and Tail Inference) *Under the assumptions of Theorem 3 or 4, as* $\tau \searrow 0, \tau T \to \infty$, $\forall l, m > 0, m \neq 1, x, \dot x \in \mathbf{X}$

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

(ii) $\hat\rho_{x,\dot x,l} - \rho_{x,\dot x,l} \xrightarrow{p} 0$, $\rho_{x,\dot x,l} \to l^{-\xi} \cdot [H(x)/H(\dot x)]$ (cf. Thm 4). *In particular,*

(iii) $\hat\xi \equiv \frac{-1}{\ln l} \ln \hat\rho_{\bar X,\bar X,l} \xrightarrow{p} \xi$,

(iv) $\hat\rho_{x,\bar X,1} \xrightarrow{p} x'\mathbf{c}$, *uniformly in* x $(\xi \neq 0)$.

(v) *for* $\pi = \mu_X' Q_H^{-1} Q_X Q_H^{-1} \mu_X$, $l = m = 2$, *if* $\sqrt{\tau T}(\rho_{\bar X,\bar X,l} - \lim_T \rho_{\bar X,\bar X,l}) \to 0$,

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

Theorem 5 is a simple corollary of Theorems 3,4 and Lemma 10. (i) is by the same steps as equation (6.34), and (i) implies (ii)-(iv), using the properties (v)-(vi) in Lemma 10. Uniformity in x in (iv) follows from the linearity of $\hat\rho_{x,\bar X,1}$. (v) follows from Theorems 3 and 4 by the delta method. The results can be strengthened to the uniform convergence in l, m, x (Chernozhukov[13]).

Theorem 5 shows that the regression quantiles spacings of the intermediate ranks consistently approximate the population spacings (result (i) and (ii)) which reveal the tail indices (results (iii) and (iv)).

Results (iii) and (v) may be especially emphasized, because the inference concerning the tail index is one of the most important problems in the statistics of extreme values, [26]. The proposed estimator $\hat\xi$ is a regression generalization of Pickand's estimator, if $l = m = 2$.[10] $\hat\xi$ consistently estimates the tail index ξ in the heteroscedastic regression Models 1 and 2. In fact, since $\mu_X' E(XX')^{-1}\mu_X = 1$ (normalize $\mu_X = (1, 0, ...)$), in

[10]Going back to Pickands, such a choice is due to practical reasons but may vary in applications.

Model 1 $\pi = 1$, so that variance equals that of Pickand's estimator in the setting with no regressors, [26]:

$$\frac{\xi^2(2^{2\xi+1}+1)}{(2(2^\xi-1)\ln 2)^2}.$$

Thus, unlike Pickand's and other unconditional estimators, $\hat{\xi}$ specifically adapts to the presence of covariates that affect the scale and shape of the conditional density. And even when covariates do not matter, there is *no* efficiency *loss* in using our estimator rather than that of Pickands.

7.2 Confidence Intervals. We offer only a brief discussion, whereas details and an empirical application can be found in [13] and [15], respectively. For brevity (and all practical purposes) assume that the tail index $\xi \neq 0$.

Resampling. Subsampling, Romano et al.[58] , is a simple, practical way of constructing the confidence intervals. The validity of subsampling is not immediate in our setting, since our statistics may be diverging. However, a simple modification brings its validity, Chernozhukov[13]. Another, different modification was proposed by Bertail et al.[8] for the ordinary sample quantiles and can be adapted here. In the empirical work, subsampling generates well-behaved, sensible confidence intervals ([15], [8]).

The nonparametric bootstrap fails in the extreme rank case. (A well known counter-example is that of extreme order statistics.) For the intermediate rank case, the bootstrap may work, at least when $a_T \to \infty$, since the statistic of interest is approximately an average. However, a simple bootstrap is unlikely to offer intervals of good quality, so that a smoothing as in Horowitz[43] may be needed. Recently Bickel and Sakov[69] have shown that for the case of sample median subsampling (with replacement) does better than the simple bootstrap. This might carry over to the intermediate rank cases.

Analytical Confidence Intervals: Intermediate Case. In the intermediate rank theory, the confidence intervals are simple to obtain. Theorem 5 provides the estimators for the tail index ξ and the index $x'\mathbf{c}$. The scaling constant a_T can be replaced by its empirical analog, see Theorems 3 and 4. This fully operationalizes the intervals. The simplicity, *convenience*, and parsimony of the limit make it a significant competitor of the central rank theory for quantiles in the range up to .25-.3, across common data sets. Chernozhukov[13] offers a monte-carlo confirmation, employing designs with different tail types, continuous and discrete covariates. These intervals out-perform the central ones (employing the methods in [50], built in S+).

Analytical Confidence Intervals: Extreme Rank Case. In this case the confidence intervals are involved but worth the trouble (Figures 2 and 5). First, approximate the distribution of $\mathbf{N}$. $\mathbf{N}$ is a Poisson Process, so its Laplace functional is

$$\Psi_{\mathbf{N}}(g) = E\exp\left[-\int g(u,x)d\mathbf{N}(u,x)\right] = \exp\left[-\int(1-e^{-g(u,x)})dm_{\xi,\mathbf{c},F_X}(u,x)\right],$$

for measurable, continuous functions g, vanishing outside the compacts sets, where $m_{\xi,\mathbf{c},F_X}(A) = E[\mathbf{N}(A)|\xi,\mathbf{c},F_x]$ is the intensity measure of $\mathbf{N}$, defined in Lemma 6. $\Psi_{\mathbf{N}}(g)$ is a continuous function of $\xi,\mathbf{c},F_X$. Thus the distribution of $\int gd\mathbf{N}$ can be consistently estimated by that of $\int gd\mathbf{N}_{\hat{\xi},\hat{\mathbf{c}},\hat{F}_X}$, where $\mathbf{N}_{\hat{\xi},\hat{\mathbf{c}},\hat{F}_X}$ is the Poisson process with the intensity measure $m_{\hat{\xi},\hat{\mathbf{c}},\hat{F}_X}$. The infinite sum $\int gd\mathbf{N}_{\hat{\xi},\hat{\mathbf{c}},\hat{F}_X}$ can be approximated by a finite sum, so that the distribution of Z_∞ can be obtained by monte-carlo,[11] as in Fig. 5. To estimate the scaling a_T, we can rely on the unconditional case, [26], [8]. Employing the approximation $\mu_X'(\beta(m\tau) - \beta(\tau)) \sim c(m^{-\xi} - 1)\tau^{-\xi}$, project $\bar{X}'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))$ on $c(m^{-\hat{\xi}} - 1)\tau^{-\hat{\xi}}$ for different τ and m to get $\hat{c}$, and set $\hat{a}_T = \hat{c}(1/T)^{-\hat{\xi}}$. If τT grows polynomially fast, $\hat{a}_T/a_T \xrightarrow{p} 1$. [This should suffice in most practical cases; in very large samples, we may further refine this as $r(m,\tau) \sim C(m^{-\xi} - 1)\tau^{-\xi}(-\ln\tau)^K$ etc.]

7.3. Which theory? Let us consider the examples in Figures 4 and 5, and look at the following factors: (i) the number of regressors, (ii) the theoretical quality of approximations, and (iii) the convenience and ease of estimating nuisance parameters.

First of all, covariates reduce the effective sample size. To that end, the concept of the effective rank is useful. The effective rank, r, is the ratio of the rank to the number of regressors, $\tau T/d$.[12] To motivate this, consider a simple "regression" quantile problem in a sample of 1000 observations and 10 dummy regressors, in which the target is the .2-th conditional quantile function. The estimate is the 20-th lowest order statistics in each of 10 subsamples corresponding to the dummy variables. Figure 2 (A-C), corresponding to this example, suggests the normal approximations are much worse than the extreme rank one. So if r is less than or equal to $25 - 40$, the Figures 2 and 5 (A-C) prefer the extreme rank approximation for the quality reasons.

When the effective rank r is above $25 - 40$, the normal intermediate or central rank approximations appear sensible. Irrespective of the sample size, the intermediate rank theory (in principle) should not be useful for the central quantiles (our modest examples suggest the range $(.3, .7)$). However, irrespective of the sample size, the intermediate rank theory is more useful for the high and low conditional quantiles ($\tau < .3, \tau > .7$) because of the simplicity of estimating nuisance parameters.

Because the intermediate rank theory exploits the extremality of the relevant events, it provides an approximation conveniently defined by the tail parameters (in contrast, the conventional theory requires the nonparametric conditional density function evaluated at the high or low quantile, which is hardly estimable with the scarce tail data and many covariates). Of course, if we fix τ and let the sample size go to infinity the central rank theory will dominate quality-wise, but the theoretical gain should be very small (Figure 5, D-F). In summary, we believe that the intermediate rank theory does better that the conventional theory at offering a more qualitative and practical approach to making inference on high and low conditional quantiles.

7.4 Other Results. In [13], we have further explored the asymptotic questions by looking at the empirical processes of the form $\left(a_T(\hat{\beta}(\tau l) - \beta(\tau l), l \in \mathcal{L}\right)$, $\mathcal{L} \subset (0, \infty)$. The convergence to either Gaussian or non-Gaussian processes have been demonstrated. These results have many practical applications in estimating tail parameters. Some hypothesis testing and refinements of tail estimators in Theorem 5 are also explored.

[11] To obtain the theoretical approximation, (a) simulate $(J_i, \mathcal{X}_i), i \leq n$; where X_i are drawn from F_X, J_i are simulated as defined in Theorems 1 and 2. (b) solve $\bar{Z}_{\infty,n} = \text{argmin} \sum_{i=1}^n \rho_{k/n}(J_i - \mathcal{X}_i'z)$; repeat (a) b times. b and n should be large. In practice, X_i may be drawn from the empirical dis-n f-n $\hat{F}_X$, and J_i are drawn as defined in Theorems 1 and 2, replacing $\xi, X_i'c$ with $\hat{\xi}, \widehat{X_i'c}$.

[12] A more refined version may be $\tau T/d$ times the determinant of the correlation matrix of X. If covariates are independent, this will give τT.

Conclusion

The present work provides a theoretical framework for studying the conditional extremal response – the near-extreme conditional quantile functions. We developed the models that coherently combine the linear forms with flexible heteroscedasticity and extreme-value-theoretic restrictions. We suggested and motivated the concepts of extremality or data scarcity for regression quantiles – the intermediate and extreme rank sequences. We obtained the limit distributions under each of these sequences.

The numerical examples in Figures 2 and 5 suggest that these distributions approximate the finite-sample distributions better or as well as the conventional theory, for quantiles in the range .01 - .3 (in samples of common size). These distributions are conveniently determined only by tail parameters. These tail parameters are easy to estimate (unlike nonparametric conditional density evaluated at near-extreme quantiles, required in the conventional theory.) We also provided the tail estimators which, unlike the widely-used Pickands and other classical procedures, specifically adapt to the setting where covariates affect location, scale, and shape of conditional density.

The relevance of these results stems from both the motivation for quantile regression models in data analysis and the importance of tail inference. The motivation was to explore many more features of the conditional distributions than just the center. For example, Abreveya[2] and Koenker and Hallock[50] characterize the economic determinants of babies' very low birth-weights through the near-extreme conditional quantiles (.05 and below). Deaton[20] examines the food expenditure of Pakistani households by the .1-th and .9 -th conditional quantiles. In our work, [15], we study the economic determinants of very high risk of an oil-producer's stock price. We find that the market factor is an unambiguously strong determinant, whereas other factors are *not*. Thus the level of extreme risks are mainly determined by the *general* economic activity. We also find and characterize the tail thickness of the *conditional* distribution, using the procedures developed here. Presently, Chernozhukov and Hong[14] are considering the robust estimation of the auction models, discussed in Example 2.4.

We shall (and hope others will) further address the inference questions, such as confidence intervals, hypothesis testing, and tail inference procedures. Over more than fifty years, an elaborate theory of inference based on the ordinary sample quantile has been developed, [26], and now forms an essence of the extreme value theory. Further progress is possible by building on these developments.

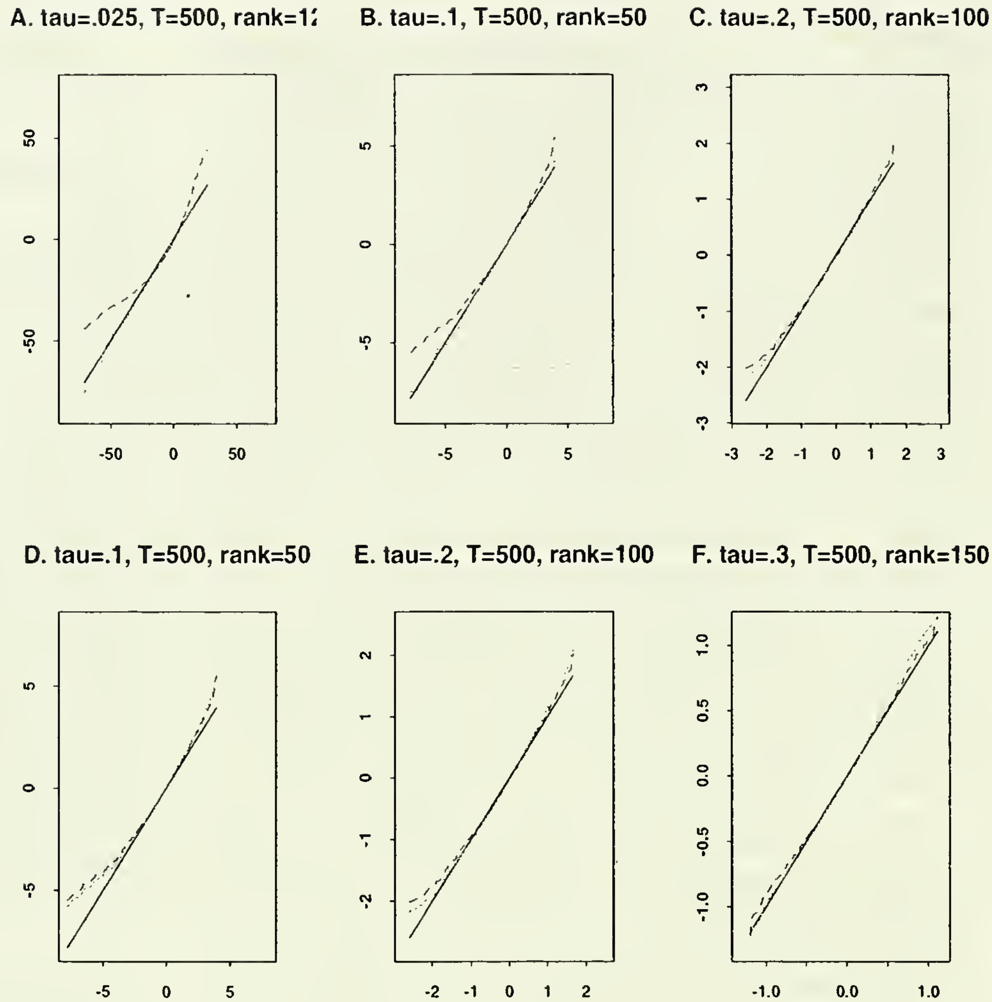

Figure 5: A simulation example for a classical model: $Y = X'\alpha + U$, U has the algebraic tail $F(u) \sim u^{-1/\xi}$ as $u \searrow \infty$, $\xi = 1$. X_1, X_2 are symmetric Beta variables (normal looking, but with bounded support), X_3 and X_4 are dummy variables. The results are for one of coefficients (others are similar). T=500. Replications = 5,000. **Displays A-C: QQ-plot of Extreme and Central Rank Approximations.** The dashed line "- - -" is the central approximation, and the dotted line "......" is the extreme rank approximation. The true quantiles of the exact sampling distribution are depicted by the solid line "——". The central rank approximation varies from very bad to bad for low quantiles $\tau = .025$ and $\tau = .1$ and becomes comparable to the extreme rank approximation only at $\tau = .2$. **Displays D-F: QQ-plot of Intermediate and Central Rank Approximations.** The dotted line "......" is now the intermediate rank approximation. The theoretical central and intermediate rank approximations have approximately the same performance for $\tau = .1, .2, .3$, (using $m = 2, 1.5, 1.25$). The practical advantage of the intermediate rank is the parsimony and ease of estimating nuisance parameters. [QQ plots are over the 99% range.]

APPENDIX. The appendix gives proofs for Theorems 1-4 in the text (Corollary 1 and 2), and studies Models 1 and 2 in detail. It first develops a set of simple "high-level" conditions, that give the requisite convergence results. These conditions are verified for Models 1 and 2, leading to fairly compact proofs of Theorems 1-4. In addition, these conditions enable an extension of the applicability domain beyond Models 1 and 2, by only supplying the key convergence results (e.g. CLTs etc). We also include some background material.

In what follows, $\{Y_t, X_t\}$ is a (triangular) sequence of random variables taking values in $\mathbb{R}^1 \times \mathbb{R}^d$ and defined on probability space $(\Omega, \mathcal{F}, P)$, *possibly indexed by* T. The outer P^* and inner P_* probability measures are defined in [76]. α-mixing or strong mixing is defined e.g. in[18]. Use $F_U(z|x)$ to denote $F_{U_t}(z|X_t = x)$ in the sequel. Unless otherwise stated, κ, K, C and their modifications are generic constants.

A Useful background

A.1 Point processes

These definitions, collected for the reader's convenience, may be found in [65] and [52].

Definition A.1 (Point Measures, $M_p(E)$) Let E be a locally compact topological space with a countable basis. Define $\mathcal{E}$ to be the Borel σ-algebra of subsets of E. A *point measure* (p.m.) p on $(E, \mathcal{E})$ is a measure of the following form: for $\{x_i, i \geq 1\}$, a countable collection of points (called points of p), and any set $A \in \mathcal{E}$:

$$p(A) \equiv \sum_i 1(x_i \in A).$$

If $p(K) < \infty$, for any $K \subset E$ compact, then p is said to be Radon. A p.m. p is simple if $p(x) \leq 1 \quad \forall x \in E$, and is compound otherwise. Let $M_p(E)$ be the collection of all Radon point measures. Sequence $\{p_n\} \subset M_p(E)$ converges vaguely to p, if $\int f dp_n \to \int f dp$ for all functions $f \in C_K^+(E)$ [continuous, non-negative, and vanishing outside a compact set] (cf. Leadbetter et.al.[52]). Vague convergence induces *vague topology* on $M_p(E)$. The topological space $M_p(E)$ is metrizable as a complete separable metric space. $M_p(E)$ denotes such a metric space hereafter. Define $\mathcal{M}_p(E)$ to be the σ-algebra generated by the open sets.

Definition A.2 (Point Processes: Convergence in Distribution.) A *point process* (PP) in $M_p(E)$ is a measurable map

$$\mathrm{N} : (\Omega, \mathcal{F}, P) \to \left(M_p\left(E\right), \mathcal{M}_p(E)\right),$$

i.e. for every elementary event $w \in \Omega$, the realization of the point process $\mathrm{N}(w)$ is some point measure in $M_p(E)$. Thus, the concept of *convergence in distribution* (in law, weak convergence) of the point process N_n taking values in $M_p(E)$ is the same as for any metric space, cf. [65]: we shall write

$$\mathrm{N}_n \Rightarrow \mathrm{N} \text{ in } M_p(E)$$

if $E_P h(\mathrm{N}_n) \to E_P h(\mathrm{N})$ [i.e. $\int_\Omega h(\mathrm{N}_n(\omega))dP(\omega) \longrightarrow \int_\Omega h(\mathrm{N}(\omega))dP(\omega)$] for all continuous and bounded functions h mapping $M_p(E)$ (or $M_+(E)$) to $\mathbb{R}$. This implies that if $\mathrm{N}_n \Rightarrow \mathrm{N}$ in $M_p(E)$,

$$\int_E f(x) d\mathrm{N}_n(x) \xrightarrow{d} \int_E f(x) d\mathrm{N}(x)$$

for any $f \in C_K(E)$ by the continuous mapping theorem.

Definition A.3 (Poisson Point Process or Random Measure (PRM)) Point process N is a PRM in $M_p(E)$ with *mean intensity measure* m defined on $(E, \mathcal{E})$, if
(a) for any $F \in \mathcal{E}$ and any non-negative integer k

$$P(\mathrm{N}(F) = k) = \begin{cases} e^{-m(F)} m(F)^k / k! & \text{if } m(F) < \infty, \\ 0 & \text{if } m(F) = \infty, \end{cases}$$

(b) for any $k \geq 1$, if $(F_i, i \leq k)$ are disjoint sets in $\mathcal{E}$, $(\mathrm{N}(F_i), i \leq k)$ are independent.

Definition A.4 (Compositions and Transformation of PRM) To construct our limit processes, the following are helpful (see Proposition 3.7 and 3.8 in Resnick[65].)
1. (Canonical PRM) The PP with points $\{\Gamma_i, i \geq 1\}$ in $M_p(E)$, where $E = [0, \infty)$, $\Gamma_i = \sum_{j \leq i} \mathcal{E}_j$, $\{\mathcal{E}_i\}$ are i.i.d. unit exponential, is PRM with mean measure $m(du) = du$ on $(E, \mathcal{E})$.
2. Let $\{V_i, i \geq 1\}$ be i.i.d. random variables with law F_V, taking values in $(S, \mathcal{S})$, satisfying definition A.1, then the PP with points $\{\mathcal{E}_i, V_i, i \geq 1\}$ is PRM in $M_p(E')$ with mean measure $m(du, dv) = du \times F_V(dv)$ on $(E', \mathcal{E}') = (E \times S, \mathcal{E} \times \mathcal{S})$.
3. Let N_1 be a PRM in $M_p(E_1)$ with points $\{\mathcal{G}_i, i \geq 1\}$ and mean measure m_1 on $(E_1, \mathcal{E}_1)$. Then the PP N_2 with points defined by $\{\mathbf{T}(\mathcal{G}_i), i \geq 1\}$, where $\mathbf{T} : (E_1, \mathcal{E}_1) \mapsto (E_2, \mathcal{E}_2)$ is measurable, is PRM in $M_p(E_2)$ with mean measure $m(dg) = m_1 \circ \mathbf{T}^{-1}(dg)$ defined on $(E_2, \mathcal{E}_2)$.

A.2 Convex Semi-Continuous Objectives

The following result is from Knight[46]. It allows for general discontinuities and $\bar{\mathbb{R}}$ - valued objective functions. The result is embedded by Knight[46] into the framework of *stochastic equi-semi-continuity* of the objective functions, which gives an elegant way of transforming the weak finite-dimensional (*fidi-*) convergence of objective functions into the weak convergence of argmins, provided the sequence of agrmins is $O_p(1)$. In case of convexity one has s.e.-sc. Related literature is [67], [70], [22].

Lemma 1 (Knight[46], p.12) *Suppose $\{Q_T\}$ is a sequence of lower-semi-continuous (lsc) convex $\bar{\mathbb{R}}$-valued random functions, defined on $\mathbb{R}^d$, and let $\mathcal{D}$ be a countable dense subset of $\mathbb{R}^d$. If Q_T fidi-converges to Q_∞ in $\bar{\mathbb{R}}$ on $\mathcal{D}$ where Q_∞ is lsc convex and finite on an open non-empty set a.s., then $\operatorname{argmin}_{z \in \mathbb{R}^d} Q_T(z) \xrightarrow{d} \operatorname{argmin}_{z \in \mathbb{R}^d} Q_\infty(z)$, provided the latter is uniquely defined a.s. in $\mathbb{R}^d$.*

B Proofs for section 5, extreme ranks

B.1 Details and definitions

Definition B.1 (Key Point Process, Space E) The key PP in $M_p(E)$ is

$$\widehat{\mathrm{N}}(\cdot) \equiv \sum_{t \le T} 1\{a_T(U_t - b_T), X_t) \in \cdot\}; \tag{B.35}$$

The mis-en-scenes 1-3 define **(1)** the measurable spaces $(E, \mathcal{E})$, **(2)** the reference error U_t, **(3)** restrictions on the constants a_T, b_T, and **(4)** the rescaled estimators

1. $E_1 \equiv [-\infty, \infty) \times \mathbf{X}$, $U_t \equiv Y_t - X_t'\beta_r$, $Z_T \equiv a_T(\hat\beta(\tau) - \beta_r - b_T \mathbf{e}_1), a_T > 0$;	2. $E_2 \equiv [[-\infty, \infty] \setminus \{0\}] \times \mathbf{X}$, $U_t \equiv Y_t - X_t'\beta_r$, $Z_T \equiv a_T(\hat\beta(\tau) - \beta_r)$, $b_T = 0,\ a_T > 0$.	3. $E_3 \equiv [0, \infty) \times \mathbf{X}$, $U_t \equiv Y_t - X_t'\beta_r \ge 0$, $Z_T \equiv a_T(\hat\beta(\tau) - \beta_r)$, $b_T = 0,\ a_T > 0$.	(B.36)

where $\beta_r = \beta(0)$ in the finite support case; $\mathbf{e}_1 = (1, 0, ..., 0)'$; $\mathbf{X}$ is a compact subset of $\mathbb{R}^d$ s.t. $X_t \in \mathbf{X}\ \forall t$. σ-algebra $\mathcal{E}$ on E is generated by the opens sets of E.

Mis-en-scene **1, 2, 3** suit the case of $F_{U_t}(\cdot|X_t)$ with **type 1** tails (finite and infinite support cases), **type 2** tails (infinite support), **type 3** tails (finite support), respectively. The scaling constants (a_T, b_T) for Models 1 and 2 are in the main text, section 3.6. They conceivably differ in other cases.

Remark B.1 (Compactification) The choice of E_1, E_2, E_3, that is their *topology*, is *important* and *simplifies the proofs* considerably. We assume that the topology on E_2 and E_1 is induced via a standard two- and one-point compactification respectively (so that e.g. $[-\infty, a] \times \mathbf{X}$ is compact *in* E_2 for $a < 0$ and *in* E_1 for any $a < \infty$.)

Definition B.2 (Limit Point Process.) We require that $\widehat{\mathrm{N}} \Rightarrow \mathrm{N}$ in $M_p(E_i)$,

$$\mathrm{N}(\cdot) = \sum_{i=1}^{\infty} 1\{(J_i, \mathcal{X}_i) \in \cdot\}, \tag{B.37}$$

where $\{J_i, \mathcal{X}_i\}$ are *random vectors* in E_i $(i = 1, 2, 3)$, finitely-valued a.s.

Definition B.3 (Normalized Statistics and Limit.) Given the definition of the quantile regression estimator, the rescaled statistic Z_T in (B.36) solves:

$$Z_T = \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ Q_T(z, k) \equiv \sum_{t=1}^{T} \rho_\tau \big(a_T(U_t - b_T) - X_t'z\big) \right\} \tag{B.38}$$

[write either $z \equiv a_T(\beta - \beta_r - b_T \mathbf{e}_1)$ or $z \equiv a_T(\beta - \beta_r)$.] The weak limit Z will solve

$$Z = \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ Q_\infty(z, k) \equiv -k\mu_X' z + \int_E l(u, x'z) d\mathrm{N}(j, x) \right\}, \tag{B.39}$$

where $l(u, v) \equiv 1(u \le v)(v - u)$, $\mu_X \equiv$ plim $\bar{X}$. Since $\mathrm{N} \in M_p(E_i)$, $Q_\infty(\cdot, k)$ is well defined and finite on an open non-empty subset of $\mathbb{R}^d$, under the conditions stated below.

Definition B.4 (Essential Uniqueness.) Let $\mathcal{K} = (K_1, K_2)$ s.t. $k \in \mathcal{K}$. Given $w \in \Omega$, let $\mathcal{K}_b(w)$, the break-points, be the set of $k_b \in \mathcal{K}$ s.t. $\arg\min_{z \in \mathbb{R}^d} Q_\infty(z, k_b)$ is not unique in $\mathbb{R}^d$. For any ω, $\mathcal{K}_b(w) \ne \emptyset$ due to the piece-wise linear form of the objective [see gradient conditions in Remark 5.1]. We require the **essential uniqueness**: either (i) $k \notin \mathcal{K}_b$ w.p. 1 or (ii) Lebesgue measure of $\mathcal{K}_b$ is zero, w.p.1.

B.2 Basic conditions for nondegenerate asymptotics

The BC are formulated as general but constructive conditions that (i) lead to short proofs of Theorems 1 and 2, (ii) allow to extend the applicability beyond Models 1 and 2, by supplying only the requisite limit laws, such as the convergence of a point process.

Condition 1 BC.1 (Key Point Process) There exist constants $\{a_T, b_T\}$ and E_i of the forms in (B.36) s.t. point process $\widehat{\mathbf{N}}$, defined in (B.35), converges weakly to $\mathbf{N}$ in $M_p(E_i)$, $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$. (For E_2 we only require $\widehat{\mathbf{N}}(\cdot \cap E_2') \Rightarrow \mathbf{N}(\cdot \cap E_2')$ in $M_p(E_2')$, where $E_2' = [-\infty, 0) \times \mathbf{X}$). Points of $\mathbf{N}$ are $\mathbb{R}^{d+1}$-valued a.s.

BC.2 (Stability and Compactness) $\{X_t\}, \{\mathcal{X}_i\}$ have the support contained in a compact set $\mathbf{X} \subset \mathbb{R}^d$. The non-degenerate limit empirical distribution functions of $\{X_t\}$ and $\{\mathcal{X}_i\}$, F_X and $F_{\mathcal{X}}$, exist with support in $\mathbf{X}$. F_X has mean μ_X.

BC.3 Design Conditions: (1) Essential uniqueness holds. (2) For space E_2, $Z_T'x < 0$ and $Z'x < 0$, $\forall x \in \mathbf{X}$, w.p. $\to 1$.

BC.4 For appropriate E_j, one of the following is true (for $\mathbf{c}(k) \in \mathbb{R}^d$): (i) $a_T(\beta(\tau) - \beta_r - b_T \mathbf{e}_1) \to \mathbf{c}(k)$, (ii) $a_T(\beta(\tau) - \beta_r) \to \mathbf{c}(k)$, (iii) $a_T(\beta(\tau) - \beta(0)) \to \mathbf{c}(k)$.

Remark B.2 **1.** BC.1 is the key. We verify it for Models 1 and 2 under dependence conditions ruling out the clusters of extremes. By supplying further convergence results, one can extend the applicability to a bigger variety of data (e.g. panels). **2.** BC.2 requires $\{X_t\}$, the properly scaled regressors, and $\{\mathcal{X}_t\}$ to have basic stability properties. Limit empirical distribution functions are known to exist under general conditions. The compactness condition is required for establishing the continuity of the mapping of $\mathbf{N}$ to the rescaled statistic. Relaxing this condition may alter the limits. BC.2 allows trends, e.g. if $X_t = t/T$, X_t has the support in $[0, 1]$, and the limiting empirical distribution is uniform. **3.** Condition BC.3, given BC.2, is plausible (see section E). **4.** Condition BC.3(2) is automatic in the main text but probably may not hold more generally. It requires that for space E_2, suiting type 2 tails, the reference line (w.l.g. the line can be chosen to be above median), is below or equal to the low rank sample regression quantiles in the compact set $\mathbf{X}$ with arbitrarily small probability as $T \to \infty$. **5.** BC.4: For Models 1 and 2, the constants $\mathbf{c}(k)$ are stated in Theorems 1 and 2 and derived in Lemma 10.

***Lemma 2 (Weak Convergence under Extreme Ranks)** BC.1-3 imply as $\tau T \to k, T \to \infty$ (for appropriate space E)*

$$\hat{Z}_T \xrightarrow{d} Z = \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ Q_\infty(z) \equiv -k\mu_X' z + \int_E l(u, x'z) d\mathbf{N}(u, x) \right\},$$

for almost every k, where $l(u, v) = 1(u \le v)(v - u)$ and $\mathbf{N}$ is defined in (B.37). This result could be stated in terms of the centered statistics. If BC.4 also holds, then

$$\text{for } Z_T^c \equiv a_T(\hat{\beta}(\tau) - \beta(\tau)), \quad Z_T^c \xrightarrow{d} Z^c \equiv Z - \mathbf{c}(k).$$

Proof: (1) Because $\tau T \to k$, take $\tau = k/T$ w.l.o.g. in sequel. First, stabilize $Q_T(z, k)$ in (B.38) by subtracting a term that does not affect optimization:

$$\widetilde{Q}_T(z, k) \equiv -k\bar{X}'z + \sum_{t=1}^{T} l_\delta(a_T(U_t - b_T), X_t'z),$$

where $l_\delta(u,v) \equiv 1(u \le v)(v-u) - 1(u \le -\delta)(-\delta - u)$, for $\delta > 0$. Such a renormalization makes the proof short, since $\widetilde{Q}_T$ now becomes a (to be shown) continuous functional of $\widehat{\mathrm{N}}$:

$$\widetilde{Q}_T(z,k) = -k\bar{X}'z + \int_{E_i} l_\delta(j, x'z) d\widehat{\mathrm{N}}(j,x). \tag{B.40}$$

Part **(2)** shows that the fidi weak limit of $\widetilde{Q}_T$ is a convex function in z:

$$\widetilde{Q}_\infty(z,k) = -k\mu_X' z + \int_{E_i} l_\delta(j, x'z) d\mathrm{N}(j,x). \tag{B.41}$$

$\widetilde{Q}_T(z,k)$ is convex $\forall T$, since it is a sum of convex functions in z. $Q_T(\cdot,k)$ is continuous, hence l.s.c. $\forall T$. By the convexity Lemma 1, for a.e. $k > 0$

$$Z_T \xrightarrow{d} Z \equiv \underset{z \in \mathbb{R}^d}{\operatorname{argmin}}\ \widetilde{Q}_\infty(z,k), \tag{B.42}$$

provided: (a) there is a non-empty open set $\mathcal{Z}_0$ s.t. $\widetilde{Q}_\infty(z,k)$ is finite a.s. $\forall z \in \mathcal{Z}_0$, and (b) Z exists and is a.s. unique for a.e. $k > 0$. (a) is shown in part 2 of this proof. (b) simply follows from BC.3(1), since $Z = \operatorname{argmin}_z Q_\infty(z,k)$, for Q_∞ defined in (B.39). Indeed, $Q_\infty(z,k)$ differs from $\widetilde{Q}_\infty(z,k)$ by $\Delta = \int_E 1(j \le -\delta)(-\delta - j) d\mathrm{N}(j,x) = \sum_{i: J_i \le -\delta}(-\delta - J_i)$, which is independent of z and $|\Delta| < \infty$ a.s. [13]

(2) Here we verify that

(i) $\widetilde{Q}_\infty(\cdot,k)$ is indeed a fidi distributional limit,

(ii) there is an open non-empty set $\mathcal{Z}_0$ s.t. $\widetilde{Q}_\infty(z,k)$ is finite a.s. for all $z \in \mathcal{Z}_0$.

(i). $\widetilde{Q}_\infty(\cdot,k)$ is a weak fidi limit of $\{\widetilde{Q}_T(\cdot,k)\}$ iff for any finite collection $(z_j, j \le l)$

$$\Big(\widetilde{Q}_T(z_j,k),\ j \le l\Big) \xrightarrow{d} \Big(\widetilde{Q}_\infty(z_j,k),\ j \le l\Big).$$

Since $\bar{X}'z_j \xrightarrow{p} \mu_X' z_j$, we only need to verify:

$$\Big(\int_{E_i} l_\delta(u, x'z_j) d\widehat{\mathrm{N}}(u,x),\ j \le l\Big) \xrightarrow{d} \Big(\int_{E_i} l_\delta(u, x'z_j) d\mathrm{N}(u,x),\ j \le l\Big). \tag{B.43}$$

Define the mapping from $M_p(E)$ to $\mathbb{R}^l$ (for $E_i = E_1, E_2', or E_3$)

$$\mathbf{T}_i : \widehat{\mathrm{N}} \mapsto \Big(\int_{E_i} l_\delta(u, x'z_j)\ d\widehat{\mathrm{N}}(u,x),\ j \le l\Big).$$

(a) Consider E_1. The map $(u,x) \mapsto l_\delta(u, x'z_j)$ is in $C_K(E_1)$, since it is uniformly continuous on E_1 by construction and vanishes outside the compact subset K *in* E_1:

$$K \equiv [-\infty, \max(\kappa, -\delta)] \times \mathbf{X}, \text{ where } \kappa = \max_{x \in \mathbf{X}, z \in \{z_1,\dots,z_l\}} x'z.$$

K is compact in E_1 since $\kappa < \infty$ by BC.2. Hence by construction $\widehat{\mathrm{N}} \mapsto \mathbf{T}_1(\widehat{\mathrm{N}})$ is continuous from $M_p(E_1)$ to $\mathbb{R}^l$. Thus $\widehat{\mathrm{N}} \Rightarrow \mathrm{N}$ in $M_p(E)$ implies $\mathbf{T}_1(\widehat{\mathrm{N}}) \xrightarrow{d} \mathbf{T}_1(\mathrm{N})$ by the continuous mapping theorem.

[13] Indeed, in case of E_3, $j \ge 0$, hence $\Delta = 0$. In case of E_1, E_2, note (i) $\mathrm{N}(K) < \infty$ a.s. for all K compact by definition of $\mathrm{N} \in M_p(E_i)$. $K = [-\infty, -\delta] \times \mathbf{X}$ is indeed a compact subset of E_2 or E_3, cf. remark B.1, so $\#\{i : J_i \le -\delta\} < \infty$ a.s., (ii) points $\{J_i\}$ of N are real-valued by *BC*.1. (i) and (ii) imply $|\Delta| < \infty$ a.s. by BC.1.

(b) Consider E_3. Map $(u,x) \mapsto l(u, x'z_j)$ is in $C_K(E_3)$: it is uniformly continuous on E_3 by construction and vanishes outside the compact subset K *in* E_3:

$$K \equiv [0, \max(\kappa, -\delta, 0)] \ \times \ \mathbf{X}, \ \text{where} \ \kappa = \max_{x \in \mathbf{X}, z \in \{z_1, ..., z_l\}} x'z.$$

K is compact in E_3 since $\kappa < \infty$ by BC.2. Therefore, by construction $\widehat{\mathbf{N}} \mapsto \mathbf{T}_3(\widehat{\mathbf{N}})$ is continuous from $M_p(E_3)$ to $\mathbb{R}^l$. Hence $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E_3)$ implies $\mathbf{T}_3(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}_3(\mathbf{N})$.

(c) Consider E_2. We claim it suffices to show the weak fidi convergence (B.43) *only* for points $\mathcal{Z}_N = \{z : x'z < 0, \forall x \in \mathbf{X}\}$, since Z_T and Z belong to such set w.p. $\to 1$, as $T \to \infty$, by BC.3. This claim is verified in Remark B.3. Note that map $(u,x) \mapsto l_\delta(u, x'z)$ is in $C_K(E_2')$ if $z \in \mathcal{Z}_N$, since it is uniformly continuous on E_2' by construction and vanishes outside the compact subset K *in* E_2':

$$K \equiv [-\infty, \max(\kappa, -\delta)] \ \times \ \mathbf{X}, \ \text{where} \ \kappa = \max_{x \in \mathbf{X}, z \in \{z_1, ..., z_l\}} x'z.$$

K is compact in E_2' since $\kappa < 0$ and $z \in \mathcal{Z}_N$. Hence $\widehat{\mathbf{N}} \mapsto \mathbf{T}_2(\widehat{\mathbf{N}})$ is continuous from $M_p(E_2')$ to $\mathbb{R}^l$. Hence $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E_2')$ implies $\mathbf{T}_2(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}_2(\mathbf{N})$.

(ii). To show (ii), pick distinct $(z_1, ..., z_l, l \geq d+1)$, so that the convex hull $\mathcal{Z}$ of these points is non-empty and has positive Lebesgue measure in $\mathbb{R}^d$; for E_2', additionally require $z_i \in \mathcal{Z}_N$ for each i (possible by compactness of $\mathbf{X}$). Define $\mathcal{Z}_0$ as the interior of $\mathcal{Z}$. By construction $\mathcal{Z}_0$ is an open, bounded, non-empty subset of $\mathbb{R}^d$. For any $z \in \mathcal{Z}_0$, $(u,x) \mapsto l_\delta(u, x'z)$ is in $C_K(E_i)$, by the arguments in (i), which implies $\int_{E_i} l_\delta(u, x'z) d\mathbf{N}(u,x)$ is finite a.s. To check this note (a) $l_\delta(u, x'z) \in C_K(E_i)$ implies $\#\{t : l_\delta(J_t, X_t'z) \neq 0\}$ is finite a.s. and (b) $l_\delta(u, x'z)$ is bounded on E_i.■

Remark B.3 Consider E_2. We claimed it suffices to show the weak fidi convergence (B.43) only for points $\mathcal{Z}_N = \{z : x'z < 0, \forall x \in \mathbf{X}\}$, since Z_T and Z necessarily belong to such set w.p. $\to 1$ by BC.3. Consider the objective functions:

$$\widetilde{Q}_T(z,k,\epsilon) \equiv \ \widetilde{Q}_T(z,k) + \phi(\sup_{x \in \mathbf{X}} x'z \leq -\epsilon),$$

$$\widetilde{Q}_\infty(z,k,\epsilon) \equiv \widetilde{Q}_\infty(z,k) + \phi(\sup_{x \in \mathbf{X}} x'z \leq -\epsilon),$$

where $\phi(A) = 0$, if A is true, $\phi(A) = \infty$ if not. They are convex and *l.s.c.* by construction, finite on an open non-empty set by the earlier arguments in part 2(ii) of the proof and by compactness of $\mathbf{X}$ (so that it is possible to choose points z s.t. $\sup_{x \in \mathbf{X}} x'z < -\epsilon$). Hence by the convexity lemma 1, $Z_T^\epsilon \equiv \ \text{argmin}_z \widetilde{Q}_T(z,k,\epsilon) \xrightarrow{d} Z^\epsilon \equiv \ \text{argmin}_z \widetilde{Q}_\infty(z,k,\epsilon)$ by the fidi weak convergence demonstrated in the proof of Lemma 1, part 2 (c); except if z_j is s.t. $\phi(\sup_{x \in \mathbf{X}} x'z_j \leq -\epsilon) = \infty$, $\widetilde{Q}_T(z_j,k,\epsilon) = \infty \to \widetilde{Q}(z_j,k,\epsilon) = \infty$ in $\bar{\mathbb{R}}$ trivially. Next choose ϵ small s.t. the probability that Z_T^ϵ and Z^ϵ differ from Z_T and Z, respectively, is as asymptotically small as desired by BC.3(2). This shows $Z_T \xrightarrow{d} Z$.

Next let $Z_T(k) \equiv \ \text{argmin}_{z \in \mathbb{R}^d} Q_T(z,k)$ and $Z(k) \equiv \ \text{argmin}_{z \in \mathbb{R}^d} Q_\infty(z,k)$.

Lemma 3 *BC.1-BC.3 imply* $\mathbf{Z}_T \equiv \left(Z_T(k_j)', j \leq l\right)' \xrightarrow{d} \mathbf{Z} \equiv \left(Z(k_j)', \ j \leq l\ \right)'$.

Proof: $\mathbf{Z}_T \in \ \text{argmin}_{\ z \in \mathbb{R}^{d \times l}} \widetilde{Q}_T(z_1,k_1) + ... + \widetilde{Q}_T(z_l,k_l)$, for $z = (z_1, ..., z_l)$. Since this objective is a sum of objective functions in Lemma 2, it retains the properties of the elements summed. Therefore the argument of Lemma 2 applies. ■

B.3 Limits in Models 1 and 2

Lemma 4, Resnick[65], states conditions (B.44) and (B.45) for convergence to a simple point process N in $M_p(E)$. Lemma 5 shows (B.45) suffices for weakly dependent data. Lemma 6 verifies (B.45) and finds the limit N in Models 1 and 2. Corollary 1 gives Theorems 1-2.

Lemma 4 (Resnick[65], 3.22) *Suppose* N *is a simple point process in* $M_p(E)$, $\mathcal{T}$ *is a basis of relatively compact open sets s.t.* $\mathcal{T}$ *is closed under finite unions and intersections, and for* $F \in \mathcal{T}$, $P(\mathrm{N}(\partial F) = 0) = 1$. *Then* $\widehat{\mathrm{N}} \Rightarrow \mathrm{N}$ *in* $M_p(E)$ *if for* $\forall F \in \mathcal{T}$:

$$\lim_{T\to\infty} P[\widehat{\mathrm{N}}(F) = 0] = P[\mathrm{N}(F) = 0], \tag{B.44}$$

$$\lim_{T\to\infty} E\widehat{\mathrm{N}}(F) = E\mathrm{N}(F) < \infty. \tag{B.45}$$

Remark B.4 In our case, $\mathcal{T}$ consists of finite unions and intersections of bounded open rectangles in E_1, E_2', E_3. Remark B.1 gives the topology of E_1, E_2', E_3.

We impose the Meyer[55] conditions on our "rare" events

$$A_t^T(F) \equiv \{w \in \Omega : (a_T(U_t - b_T), X_t) \in F\}.$$

Lemma 5 (Poisson Limits) *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 the mixing coefficient* $\alpha_T(\cdot)$, *(B.45) holds, and there exist sequences of integers* $(p_n, n \ge 1)$, $(q_n, n \ge 1)$, $(t_n = n(p_n + q_n), n \ge 1)$: *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) that* $I_{p_n} = \sum_{i=1}^{p_n-1}(p_n - i)P(A_1^{t_n}(F) \cap A_{i+1}^{t_n}(F)) = o(1/n)$. *Then in* $M_p(E)$, $\widehat{\mathrm{N}} \Rightarrow \mathrm{N}$, *a PRM with mean measure* $m : m(F) \equiv \lim_{T\to\infty} E\widehat{\mathrm{N}}(F)$.

Proof. For any $F : m(F) > 0$, $\lim_{T\to\infty} P[\widehat{\mathrm{N}}(F) = 0] = P[\mathrm{N}(F) = 0] = e^{-m(F)}$, by Meyer's theorem[55]. The same also holds for $F : m(F) = 0$, since $E\widehat{\mathrm{N}}(F) \to 0$ implies $P(\mathrm{N}(F) = 0) \to 1$ [$\widehat{\mathrm{N}}(F)$ is integer-valued]. Conclude by lemma 4 and definition of PRM. ■

Remark B.5 $I_{P_n} = o(1/n)$ prevents clusters of "rare" events $A_t^T(F)$. It eliminates compound Poisson processes as limits. The Meyer condition generalizes Loynes[53]. Leadbetter et al[52] offer generalization, distributional mixing, not suited when we have X.

Lemma 6 (Limit N in Models 1 and 2) *Under Assumption 2 or 1, and dependence conditions in lemma 5, for the canonical constants* (a_T, b_T) *defined in terms of* F_u *in section 3.6,* $\widehat{\mathrm{N}} \Rightarrow \mathrm{N}$ *in* $M_p(E_i)$, *a PRM with mean measure defined on* $E(E_1, E_2', or E_3)$ *as:*

$$m(du, dx) = K(x)h(du) \times F_X(dx),$$

where $h(u) = e^u$ *for type 1,* $h(u) = (-u)^{-\alpha}$ *for type 2, and* $h(u) = u^\alpha$ *for type 3 tails. Points* $(J_i, \mathcal{X}_i)$ *of* 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, $\Gamma_i = \mathcal{E}_1 + ... + \mathcal{E}_i, i \ge 1$ *(*$\{\mathcal{E}_i\}$ *are i.i.d. standard exponential), and* $\{\mathcal{X}_i\}$ *are i.i.d. r.v. with law* F_X, *independently distributed from* $\{\Gamma_i\}$.

In view of the form of $K(\cdot)$ (Lemma 10, see assumption 2), the points of N are

$$(J_i, \mathcal{X}_i, i \ge 1) \stackrel{d}{=} \begin{cases} (\ln(\Gamma_i) + \mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 1,} \\ (-\Gamma_i^{-1/\alpha}\mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 2,} \quad i \ge 1 \\ (\Gamma_i^{1/\alpha}\mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 3.} \end{cases} \tag{B.46}$$

Proof of Lemma 6: By Lemma 4 and 5, this reduces to verifying that the mean intensity measure $E\widehat{\mathbf{N}}(F)$ converges to $m(F)$, $\lim_T E\widehat{\mathbf{N}}(F) = m(F)$ for all F in $\mathcal{T}$, which follows by straightforward calculations. To show the second part, construct a PRM with the intensity measure $m(\cdot)$. At first, define a canonical homogeneous PRM $\mathbf{N}_1$ with points $\{\Gamma_i, i \geq 1\}$ (defined as above). It has the mean measure $m_1(du) = du$ on $[0, \infty)$, e.g. Resnick [65]. Secondly, by the (Composition) Proposition 3.8 in [65], the *composed* PP $\mathbf{N}_2$ with points $\{\Gamma_i, \mathcal{X}_i\}$ is PRM with the mean measure

$$m_2(du, dx) = du \times F_X(dx)$$

on $[0, \infty) \times \mathbf{X}$, because $\{\mathcal{X}_i\}$ are i.i.d. and are independent of $\{\Gamma_i\}$ (see Def. A.4). Finally, the PP $\mathbf{N}$ 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 $E \times \mathbf{X}$

$$m(dj, dx) = m_2 \circ \mathbf{T}^{-1}(dj, dx) = K(x)h(dj) \times F_X(dx),$$

by the Transformation Proposition 3.7 in Resnick[65] (see Def. A.4). ■

Corollary 1 (Theorems 1-2) *Lemma 3-6, Lemma 10, and section D verified the conditions BC.1-4 (E_1, E_2, E_3 suit the tail types 1-3.) Hence we have Theorems 1-2.*

C Proofs for section 6, intermediate ranks

C.1 Basic conditions for a normal limit

Define the following key variables:

$$W_T(l) \equiv \frac{-1}{\sqrt{(\tau l)T}} \sum_t \left(l\tau - 1(Y_t \leq X_t'\beta(\tau l))\right) X_t,$$

$$G_T(l, z) \equiv \frac{1}{\sqrt{(\tau l)T}} \sum_t \mu_t(l, z) \left[-X_t'z + (Y_t - X_t\beta(\tau l))\, a_T(l)\right],$$

where $\mu_t(l, z) = \left(1(Y_t \leq X_t'\beta(\tau l)) - 1(Y_t \leq X_t'\beta(\tau l) + X_t'z/a_T(l))\right)$.

Condition 2 *There exists a sequence $\{a_T(l)\}$ such that as $\tau T \to \infty, \tau \searrow 0$*
*BC*1 (Analytical Tail Property) $\lim_T E_P G_T(l, z) = \frac{1}{2}z'J(l)z$, $J(l)$ is invertible $\forall l > 0$.*

*BC*2 (LLN) $\left\|G_T(l, z) - E_P G_T(l, z)\right\| \xrightarrow{p} 0$ for any fixed l and z.*

*BC*3 (CLT) $\{W_T(l_1), \ldots W_T(l_m)\} \xrightarrow{d} N(0, \mathcal{G})$ for any finite collection $0 < l_i, i \leq m$.*

Remark C.1 **1.** The (properly scaled) objective function is $W_T(l)'z + G_T(l, z)$. The BC imply its linear-quadratic normal limit, as in the main text. **2.** The analytical condition BC*1 is **most important**. Under *independence* it implies BC*2,while BC*3 holds by Lindeberg-Feller. To check, write $G_T(z, l) \equiv \frac{1}{\sqrt{\tau T}} \sum_t r_t(z, l)$, BC*1 requires $Er_t(z, l) \sim \sqrt{\tau T^{-1}}c$. By compactness of $\mathbf{X}$ and binomiality of r_t (a binomial variable times a bounded variable): $\mathrm{var}(\sum_{t=1}^T r_t(z, l)/\sqrt{\tau T}) = O(Er_t(z, l)/\tau) = O(1/\sqrt{\tau T}) \to 0$. BC*1 is *purely* analytical. Section C.2 verifies BC*1 in Models 1 and 2 under the additional assumption 3. **3.** Many CLT/LLN carefully imply BC*1 and BC*2 (e.g. [18], [24], [32]). *Carefulness* means CLTs should require no more than two asymptotically bounded moments of $W_T(l)$. E.g. Liapunov, L_2 mixingale, and NED CLTs don't apply; but Lindeberg and L_1-mixingale CLTs do. In Lemma 9 we adapt the local CLTs in Robinson[66], designed for kernel estimators.

Lemma 7 (Weak Convergence under Intermediate Ranks) BC*1- BC*3 imply, as $\tau T \to \infty, \tau \searrow 0$: $\left(a_T(l_i)(\hat{\beta}(l_i\tau) - \beta(l_i\tau)), i \le m\right) \xrightarrow{d} N(0, \Omega)$, $\Omega_{ij} = J^{-1}(l_i)\mathcal{G}_{ij}J^{-1}(l_j)$.

Proof. Case of $m = 1$ is notationally identical to the proof stated after Theorem 3 in the main text, which required only the conditions above. The (joint convergence) proof for $m \ge 2$ is very analogous, so the undue repetition is avoided. ■

C.2 Limits in Models 1 and 2

This section verifies that the conditions BC*1 and BC*2 hold in Model 2 (and 1)

Lemma 8 (Analytical Tail Property) Under assumptions 1 or 2 and 3, BC*1 holds, with $J(l)$ given in Theorems 4 and 3 for Models 2 and 1, respectively.

Proof: Suppress l ($l = 1$). Write $E(G_T(z)) = \sum_{t=1}^T \frac{E\mu_t X_t' z}{\sqrt{\tau T}} + \frac{E\eta_t}{\sqrt{\tau T}}$, where $\eta_t \equiv \mu_t(Y_t - X_t'\beta(\tau))a_T$. Use F_t and f_t to denote $F_{U_t}(\cdot|X_t)$ and $f_{U_t}(\cdot|X_t)$:

$$
\begin{aligned}
\frac{E\mu_t X_t' z}{\sqrt{\tau T}} &\equiv \frac{1}{\sqrt{\tau T}} E|F_t[F_t^{-1}(\tau)] - F_t[F_t^{-1}(\tau) + \frac{X_t' z}{a_T}]| \cdot |X_t' z| \\
&\overset{(1)}{\equiv} \frac{1}{\sqrt{\tau T}} E|\frac{f_t\{F_t^{-1}(\tau) + o\left(F_u^{-1}(m\tau) - F_u^{-1}(\tau)\right)\}}{a_T}| \cdot (z'X_t)^2 \\
&\overset{(2)}{\sim} \frac{1}{\sqrt{\tau T}} E|\frac{f_t(F_t^{-1}(\tau))}{a_T}| \cdot (z'X_t)^2 \qquad \text{(C.47)} \\
&\equiv \frac{1}{T} E|\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_t[F_t^{-1}(\tau)])^{-1}}| \cdot (z'X_t)^2 \\
&\overset{(3)}{\sim} \frac{1}{T} E\frac{1}{H(X_t)}\left|\frac{m^{-\xi} - 1}{-\xi}\right| \cdot z'X_t X_t' z, \text{ uniformly in } t.
\end{aligned}
$$

Equality (1) is from the definition of $1/a_T = o(F_u^{-1}(m\tau) - F_u^{-1}(\tau))$ and a Taylor expansion. The equivalence (2) is by the assumed (Assumption 3) regular variation and uniform in t tail-equivalence: $1/f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau \in \mathcal{R}_{-\xi-1}$. Pick $x = \mu_X$ so that $K(x) = 1$ for now. By the definition of regular variation, *locally uniformly* in l [uniformly in l in any compact subset of $(0, \infty)$]

$$f_u(F_u^{-1}(l\tau)) \sim l^{\xi+1} f_u(F_u^{-1}(\tau)). \qquad \text{(C.48)}$$

I.e. locally uniformly in l

$$f_u(F_u^{-1}(\tau) + [F_u^{-1}(l\tau) - F_u^{-1}(\tau)]) \sim l^{\xi+1} f_u(F_u^{-1}(\tau)). \qquad \text{(C.49)}$$

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

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

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

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

because for any such $\{v_\tau\}$ we can find a sequence $l_\tau \to 1$ s.t. $\{v_\tau\} = \{[F_u^{-1}(l_\tau\tau) - F_u^{-1}(\tau)]\}$. Now because (a): $1/f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau$ uniformly in t by Assumption 3, and (b): $f_u(F_u^{-1}(l\tau/K)) \sim (l/K)^{\xi+1} f_u(F_u^{-1}(\tau))$, locally uniformly in l and uniformly in $K \in K_{\mathbf{X}} = \{K(x) : x \in \mathbf{X}\}$ [compact by assumptions on $K(\cdot)$ and $\mathbf{X}$], by (C.48); the equivalence (2) in (C.47) now follows uniformly in t.

The equivalence (3) in (C.47) follows from(C.50)-(C.52). At first, by assumption $f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau \equiv 1/\{K(x)f_u[F^{-1}(\tau/K(x))]\}$, uniformly in t. Hence, uniformly in $x = X_t$:

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_t[F_t^{-1}(\tau)])^{-1}} \sim \frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(K(x)f_u[F_u^{-1}(\tau/K(x))])^{-1}}. \tag{C.50}$$

But by Lemma 10(v) and Fact D.1(ii), uniformly in $x \in \mathrm{X}$:

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau/K(x)) - F_u^{-1}(\tau/K(x))} \to 1/H(x), \tag{C.51}$$

where $H(x) = 1$ if $\xi = 0$, $H(x) = x'\mathbf{c}$ if $\xi \neq 0$. And

$$\begin{aligned}\frac{F_u^{-1}(m\tau/K(x)) - F_u^{-1}(\tau/K(x))}{\tau(K(x)f_u[F_u^{-1}(\tau/K(x))])^{-1}} &\equiv \int_1^m \frac{f_u[F_u^{-1}(\tau/K(x))]}{f_u[F_u^{-1}(s\tau/K(x))]}ds \\ &\overset{(1)}{\sim} \int_1^m s^{-\xi-1}dls = \frac{m^{-\xi}-1}{-\xi},\end{aligned} \tag{C.52}$$

where the equivalence (1) is by the assumed regular variation property (C.48).

Finally, calculations for the term $\frac{E\eta_t}{\sqrt{\tau T}}$ are analogous to (C.47), using change of variables:

$$\frac{E\eta_t}{\sqrt{\tau T}} \sim \frac{-1}{2T}E\frac{1}{H(X_t)}\left|\frac{m^{-\xi}-1}{-\xi}\right|z'X_tX_t'z, \text{ uniformly in } t \tag{C.53}$$

Combine (C.53) and (C.47) to conclude. ■

Lemma 9 (CLT & LLN) *Assume Model 2 and Assumption 3. Let $\{Y_j, X_j\}_{-\infty}^t$ be a stationary α-mixing triangular sequence. (i) If $\alpha_j = O(j^{-\phi})$, $\phi > 2$, and for any K sufficiently close to* 0^+ or $-\infty$, *uniformly in t and $s \geq 1$, there is $C > 0$, independent of K s.t. (P_t denotes $P(\cdot|\mathcal{F}_t), \mathcal{F}_t \equiv \sigma(\{Y_j, X_j\}_{-\infty}^{t-1})$):*

$$P_t(U_t \leq K, U_{t+s} \leq K) \leq CP_t(U_t \leq K)^2, \tag{C.54}$$

*BC*3 holds with $\mathcal{G}_{ij} \equiv \mathcal{Q}_X \min(l_i, l_j)/\sqrt{l_i l_j}$. If (C.54) is dropped, BC*3 still holds with $\mathcal{G}_{ij} \equiv \lim_T Ep(l_i)p(l_j)$, $p(l) \equiv s_1 - E(s_1|\mathcal{F}_1)$, $s_t \equiv (1[Y_t \leq X_t'\beta(l\tau)] - \tau l)X_t/\sqrt{\tau l}$. (ii) If $\alpha_j = O(j^{-\phi}), \phi > \frac{1}{1-\gamma}$, $0 < \gamma < 1$, and $\tau^{1-\frac{2}{\gamma}}/T \to 0$, then BC*2 holds.*

Remark C.2 (C.54) means the extremal events should not cluster too much. It may be refined slightly along Watts et. al.[77](no-cov case). (C.54) is analogous to the local no-clustering conditions of Robinson[66] (A7.4., p.191) in the context of kernel density estimation.

Proof of Lemma 9: (i) $\{W_T(l_i), i \leq m\}$ suits the CLT of Robinson[66]. His condition A7.1 (with $q = 0$), A7.2, and A7.3. are satisfied automatically. The assumed mixing condition implies $\sum_{j=1}^\infty j\alpha_j < \infty$, which implies his condition A3.3. Lastly, condition (C.54) insures A7.4. If (C.54) does not hold, apply theorem of M.I. Gordin (see [39], p.137). It is easy to check, using the assumed condition and the classical Ibragimov inequality [18], that $\{s_t, \mathcal{F}_t\}$ is stationary L_1-mixingale of size -1, and $E\|W_T(l)\|_2 < K$, uniformly in T, thus $E\|W_T(l)\|_1 < K'$. This verified the conditions of the Gordin theorem. (ii) Suppress l in notation (irrelevant). $Var(G_T(z)) = \tau^{-1}O(Var(\mu_1) + 2\sum_{k=1}^T E\mu_1\mu_{1+k})$, for μ_t defined earlier. By binomiality of μ_t and the calculations analogous to those in Lemma 8 (denote by $\|\cdot\|_{r,P}$ the $L_r(P)$ norm):

$$Var(\mu_t) = O(\|\mu_t\|_{1,P}) = O(f_u(F_u^{-1}(\tau))a_T^{-1}) = O(\sqrt{\tau/T}),$$

$$\|(\mu_1\mu_{1+s})\|_{1,P} = O(\alpha_s^{1-\gamma}\|\mu_1\|_{r,P}\|\mu_1\|_{p,P}) = O(\alpha_s^{1-\gamma}\|\mu_1\|_{1,P}^\gamma) = O\left(\alpha_s^{1-\gamma}(\tfrac{\tau}{T})^{\gamma/2}\right)$$

(for $1/p + 1/r = \gamma \in (0,1)$, $p \geq 1$), by Ibragimov ineq-ty ([18]). So $Var(G_T(z))=o(1)$. ■

Corollary 2 (Theorem 3 and 4) *Lemmas 7, 8, 9 yield theorems 3 and 4.*

D Properties of Models 1 and 2

This section derives the essential properties of Models 1 and 2. Denote by M any compact subset of $(0,\infty)$ that does not contain 1. Let $\mathcal{T}$ be the set of quantile indices $\{\tau : \tau = s\tau', s \in \mathcal{L}\}$, where $\mathcal{L}$ is any fixed compact subset of $(0,\infty)$, and τ' is the reference index $\searrow 0$.

Lemma 10 (Properties) *Assumption 2 and linearity $[Q_{Y_t}(\tau|X_t = x) = x'\beta(\tau)]$ imply*

(i) $K(x)$ is a function of linear index $K(x'\mathbf{c})$ defined after Assumption 2, section 4.

(ii) Centering constants $\mathbf{c}(k)$ are those stated in Theorem 2. (Theorem 1 for Model 1).

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

(iii) For $\mu \equiv \mathbf{c}_{-1}/(m^{-\xi} - 1)$ if $\xi \neq 0$, $\mu \equiv \mathbf{c}_{-1}/\ln m$ if $\xi = 0$,

$$\beta_1(\tau) - \beta_{1r} \sim F_u^{-1}(\tau), \tag{D.55}$$

$$\beta_{-1}(\tau) - \beta_{-1r} \sim \mu[F_u^{-1}(m\tau) - F_u^{-1}(\tau)], \tag{D.56}$$

also if $\xi \neq 0$ $\beta_{-1}(\tau) - \beta_{-1r} \sim \mathbf{c}_{-1}F_u^{-1}(\tau)$;

$$(vi) \quad \frac{(x-\mu_X)'(\beta(\tau) - \beta_r)}{\mu_X'(\beta(m\tau) - \beta(\tau))} \to \begin{cases} (x-\mu_X)'\frac{\mathbf{c}}{m^{-\xi}-1} & \text{if } \xi \neq 0, \\ (x-\mu_X)'\frac{\mathbf{c}}{\ln m} & \text{if } \xi = 0, \end{cases} \tag{D.57}$$

$$(v) \quad \frac{x'(\beta(m\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \to \begin{cases} x'\mathbf{c} & \text{if } \xi \neq 0, \\ 1 & \text{if } \xi = 0, \end{cases} \tag{D.58}$$

$$(vi) \quad \frac{x'(\beta(m\tau) - \beta(\tau))}{x'(\beta(l\tau) - \beta(\tau))} \to \frac{m^{-\xi} - 1}{l^{-\xi} - 1}. \tag{D.59}$$

We need a few facts. Note that these are for *low*, not *high*, quantiles. Write $F_u \in D(H_\xi)$, if F_u is a cdf in the domain of minimum attraction with the tail index ξ (Def. 3.2). Write $F_u \in \mathcal{R}_\gamma(0)$, if F_u is a regularly varying function with exponent γ at 0.

Fact D.1 (On Regular Variation) *Uniformly in $(m, l, \tau) \in M \times M \times \mathcal{T}$, as $\tau' \searrow 0$,*

(i) If $F_1(z) \sim F_2(z)$ as $z \searrow 0$ or $-\infty$ and $F_1 \in D(H_\xi)$, then: $F_2 \in D(H_\xi)$, $F_1^{-1}(\tau) \sim F_2^{-1}(\tau)$, $F_1^{-1}(m\tau) - F_1^{-1}(\tau) \sim F_2^{-1}(m\tau) - F_2^{-1}(\tau)$, and $F_1^{-1}, F_2^{-1} \in \mathcal{R}_{-\xi}(0)$.

(ii) Suppose $F_u(z|x) \sim K(x)F_u(z)$ as $z \searrow 0$ or $-\infty$ uniformly in $x \in \mathbf{X}$(compact) as $z \searrow 0$ or $-\infty$, and $K(\cdot) > 0$ is continuous and uniformly bounded away from zero and above. Then $F_u^{-1}(\tau|x)$ relates to $F_u^{-1}(\tau/K(x))$ as in (i) uniformly in $x \in \mathbf{X}$.

(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 $F_u \in D(H_\xi)$.

(iv) $\frac{F_u^{-1}(lm\tau) - F_u^{-1}(l\tau)}{\ell(F_u^{-1}(\tau))} \to \ln m$ if $F_u \in D(H_0)$, where ℓ is auxiliary function in section 3.6.

Except for (ii), (i)-(iv) are found in the texts on extreme values – [19], also [65], [52], [26]. (ii) holds from (i) pointwise, and uniformly – by linearity of $F_{u_t}^{-1}(\tau|x)$ in x and compactness of $\mathbf{X}$.

Proof of Lemma 10. Here 'locally uniformly' means 'uniformly in $(l, m, \tau, x) \in M \times M \times \mathcal{T} \times \mathbf{X}$, as $\tau' \searrow 0$'. Recall also $(x - \mu_X)_1 = 0$ by assumption (X includes the intercept).

First, note $\mu_X'(\beta(\tau) - \beta_r) \equiv \beta_1(\tau) - \beta_{1r} = F_u^{-1}(\tau|\mu_X) \sim F_u^{-1}(\tau)$ by assumption. And

$$\begin{aligned} \frac{(x-\mu_X)'(\beta(\tau) - \beta_r)}{\mu_X'(\beta(m\tau) - \beta(\tau))} &\equiv (x-\mu_X)'A(\tau) \equiv \frac{F_u^{-1}(\tau|x) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \\ &\sim \frac{F_u^{-1}(\tau/K(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \text{ locally uniformly} \\ &\sim \frac{(1/K(x))^{-\xi} - 1}{m^{-\xi} - 1} \text{ locally uniformly} \quad \equiv B(x), \end{aligned} \tag{D.60}$$

which follows from the facts (ii) and (iii). Since $0 < B(x) < \infty$, (D.60) implies

$$\left|(x-\mu_X)'A(\tau) - B(x)\right| \to 0 \text{ locally uniformly,} \tag{D.61}$$

and that, since $(x-\mu_X)_{-1}$ ranges over a non-degenerate subset of $\mathbb{R}^{d-1}$,

$$A_{-1}(\tau) \to \kappa(m) \text{ locally uniformly,} \tag{D.62}$$

where $\kappa(m)$ is a constant vector. Hence $B(x)$ is *affine* in x on $\mathbf{X}$.

If $\xi = 0$, conclude $B(x) = -\ln K(x)/\ln m = \mathbf{c}(x-\mu_X)/\ln m$, that is $K(x) = e^{-x'\mathbf{c}}$ on $\mathbf{X}$ where $\mu_X'\mathbf{c} = 0$, i.e. $\mathbf{c}_1 = 0$. If $\xi \neq 0$, conclude $B(x) = (K(x)^{\xi} - 1)/(m^{-\xi} - 1) = \mathbf{c}(x-\mu_X)/(m^{-\xi}-1)$, that is $K(x) = (x'\mathbf{c})^{1/\xi}$ on $\mathbf{X}$, where $\mu_X'\mathbf{c} = 1$, $\mathbf{c}_1 = 1$ $(\mu_X = (1,0,...))$.

By the assumption on $K(\cdot)$, $x'\mathbf{c}$ is uniformly bounded on $\mathbf{X}$; for types 2 and 3, $x'\mathbf{c}$ is also uniformly positive on $\mathbf{X}$. This shows claim (i).

Claim (iv) is verified by substituting the forms of $K(x)$ found into (D.60).

Claim (iii) follows directly from (D.60), (D.62), and the preceding paragraph. [Note also: $F_u^{-1}(m\tau) - F_u^{-1}(\tau) \sim (m^{-\xi}-1)F_u^{-1}(\tau)$ if $\xi \neq 0$, locally uniformly].

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

$$a_T(\beta(\frac{k}{T}) - \beta_r) \sim a_T\mathbf{c}F^{-1}(\frac{k}{T}) = \mathbf{c}F_u^{-1}(\frac{k}{T})/F_u^{-1}(\frac{1}{T}) \sim k^{-\xi}\mathbf{c}, \tag{D.63}$$

since by the fact (i) $F_u^{-1} \in \mathcal{R}_{-\xi}(0)$. If $\xi = 0$, by claim (iii), facts (i) and (iv), and the definition of vector $\mathbf{c}$ ($\mathbf{c}_1 = 0$), uniformly in k in any compact subset of $(0,\infty)$ as $T \to \infty$

$$\begin{aligned}
&a_T(\beta(\frac{k}{T}) - \beta_r - b_T\mathbf{e}_1) \sim \\
&\frac{1}{\ell(F^{-1}(1/T))}\left[\mathbf{c}\left(F_u^{-1}(e\frac{k}{T}) - F_u^{-1}(\frac{k}{T})\right) + \mathbf{e}_1\left(F_u^{-1}(\frac{k}{T}) - F_u^{-1}(\frac{1}{T})\right)\right] \\
&\to \mathbf{c}\ln e + \mathbf{e}_1 \ln k = \mathbf{c} + \mathbf{e}_1 \ln k.
\end{aligned} \tag{D.64}$$

Claim (v) holds pointwise in x by facts (ii) and (iii). Since the ratio on the l.h.s. in (D.58) is linear in x and $\mathbf{X}$ is compact, it also holds uniformly in $x \in \mathbf{X}$.

Finally, combine fact (iii) with claim (v) to have claim (vi). ■

E Design for Extreme Ranks

This design condition insures the essential uniqueness needed for Theorems 1 and 2.

Condition BC.3• (Sufficiency for Uniqueness & $O_p(1)$)

(a) For any set $\mathbb{X}$ in $\mathbf{X}$ s.t. $\int_{\mathbb{X}} dF_X > 0$ and any $\kappa_1, \epsilon > 0$, there is κ_2 large s.t.

$$\mathbf{N}([-\infty,\kappa_2] \times \mathbb{X} \cap E) > \kappa_1 \text{ w. pr. } \geq 1 - \epsilon\ . \tag{E.65}$$

(b) $F_{\mathcal{X}}$ has an absolutely continuous component (if $d \geq 2$), and $\{J_i\}$ are continuously distributed cond'l on $\{\mathcal{X}_i\}$. $F_{\mathcal{X}}$ is non-degenerate in $\mathbb{R}^d$. Normalize $\mu_X = (1,0,...)'$.

Remark E.1 Assumptions (a) and (b) trivially hold for all of the limit Poisson RM obtained in Lemma 6 (for Tms 1 and 2). One continuous covariate insures a.s. uniqueness.

Fact E.1 (About Nondegeneracy) *Nondegeneracy of distribution function F with support $S \subset \mathbb{R}^d$ means that for some positive constants δ and K $\exists$ sets R_i, $i \leq K$ that cover S s.t. for any $c: \|c\|_2 > 1$, $\exists i: \int_{R_i} dF > 0$ and $x'c > \delta\|c\|_2$ for all $x \in R_i$; cf.[60].*

Lemma 11 (Essential Uniqueness and Tightness) Under BC.3$^\bullet$, for $\mathcal{K} = [K_1, K_2]$
(i) $k \notin \mathcal{K}_b$ w.p.1. if $d \geq 2$, and $Leb(\mathcal{K}_b) = 0$, if $d = 1$ (no-covariate case),
(ii) $\sup_{k\in\mathcal{K}} \|Z(k)\| = O_{P^*}(1)$.

Proof of (i) The assumption (i) and the gradient conditions for (B.39) in the Remark 5.1 in section 5 [applicable given the tightness (ii)] insure that for any given k, $k \in \mathcal{K}_b$ w.p. 0, if $d \geq 2$. When $d = 1$, $\mathcal{K}_b$ consists of integers $\{1, 2, ...\} \cap \mathcal{K}$. ■

Proof of (ii) Select $z^f \in \mathbb{R}^d$ s.t.

$$\sup_{k\in\mathcal{K}} \left| Q_\infty(z^f, k) \equiv -k\mu_X' z^f + \int_E l(u, x'z^f) d\mathrm{N}(u, x) \right| = O_{P^*}(1). \tag{E.66}$$

(E.66) is possible as shown in the Proof of Theorem 2 and because $k\mu_X$ enters Q_∞ linearly. By the linearity and convexity, if z_1 and z_2 are s.t. (E.66) holds, (E.66) also holds for any z_3 in the convex hull of z_1, z_2.

Consider ball $B(M)$ with radius M, centered at z^f, and let $z(k) = z^f + \delta(k)v(k)$, where $v(k)$ is a unit direction vector s.t. $\|v(k)\|_2 = 1$, and $\delta(k) > M$ for any k. k and $v(k)$ *are not fixed*. By convexity in z, for all $k \in \mathcal{K}$

$$\frac{M}{\delta(k)} \big(Q_\infty(z(k), k) - Q_\infty(z^f, k) \big) \geq Q_\infty(z^*(k), k) - Q_\infty(z^f, k), \tag{E.67}$$

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

$$P_* \big(\inf_{k\in\mathcal{K}} Q_\infty(z^*(k), k) > K \big) \geq 1 - \epsilon. \tag{E.68}$$

(E.68), combined with (E.66), implies r.h.s. of $(E.67) > C > 0$ w.p. arbitrarily close to 1 for M large enough, which verifies claim (ii) of the lemma.

Thus it remains to verify (E.68). For any direction $v(k)$, as $M \to \infty$,

- (a) $\mu_X' z^*(k) = z_1^f(k) + v_1(k) \cdot M, \quad 1 \geq v_1(k) \geq 0,$
- (b) $\mu_X' z^*(k) = z_1^f(k) + v_1(k) \cdot M, \ -1 \leq v_1(k) < 0.$

(Cases (a) are the worst). Fix some $\kappa_1 > 1$. In view of (a) and (b), because $v_1(k) \leq 1$, it suffices to show that: for any K and $\epsilon > 0$, uniformly in $k \in \mathcal{K}$

$$-\kappa_1 M + \int_E l(u, x'z^*(k)) d\mathrm{N}(u, x) > K \text{ w. pr. } \geq 1 - \epsilon \text{ , as } M \to \infty, \tag{E.69}$$

and therefore (E.68). Hence it suffices to show that uniformly in $k \in \mathcal{K}$ for some $\kappa_2 > \kappa_1$

$$\int_E l(u, x'z^*(k)) d\mathrm{N}(u, x) > \kappa_2 M - \kappa_3 \text{ w. pr. } \geq 1 - \epsilon \text{ , as } M \to \infty, \tag{E.70}$$

where κ_3 is some constant. By Fact E.1, for any direction $v(k) : \|v(k)\|_2 = 1, k \in \mathcal{K}$, there is $\mathbb{X}_{i(k,v)} = \{x \in \mathbf{X} : x'z^*(k) \geq \kappa_4 M\}$ s.t. $\int_{\mathbb{X}} dF_X > 0$, $\kappa_4 > 0$, and at most K such sets $\mathbb{X}_{i(k,v)}$ correspond to all possible directions $v(k) : \|v(k)\|_2 = 1$ and $k \in \mathcal{K}$. Hence for any $(z^*(k), k)$ as $M \to \infty$

$$\begin{aligned} \int_E l(u, x'z^*(k)) d\mathrm{N}(u, x) &\geq \int_{E\cap[-\infty,\kappa_5]\times\mathbb{X}_{i(k,v)}} (x'z^*(k) - u)^+ d\mathrm{N}(u, x) \\ &\geq \mathrm{N}(E \cap [-\infty, \kappa_5] \times \mathbb{X}_{i(k,v)})(\kappa_4 M - \kappa_5)^+. \end{aligned} \tag{E.71}$$

By BC(3$^\bullet$) (a), we can select κ_5 large enough s.t. $\mathrm{N}(E \cap [-\infty, \kappa_5] \times \mathbb{X}_i) > \kappa_2/\kappa_4$ for all $i \leq K$ w. pr. $\geq 1 - \epsilon$. Now let $M \to \infty$. ■

References

[1] A. Abadie, J. Angrist, and G. Imbens, *Iv estimates of the effects of subsidized training on the quantiles of trainee earnings*, forthcoming in Econometrica, 2001.

[2] J. Abreveya, *The effect of demographics and maternal behavior on the distribution of birth outcomes*, Empirical Economics (2001), forthcoming.

[3] V. Aguirregabiria, *The dynamics of markups and inventories in retailing firms*, The Review of Economic Studies **66** (1999), 275–308.

[4] J. Altonji and H. Ichimura, *Estimation of nonseparable econometric models*, working paper, Northwestern, 1999.

[5] K. Arrow, T. Harris, and J. Marschak, *Optimal Inventory Policy*, Econ-ca **19** (1951), 205–272.

[6] P. Bajari, *Auction models are not robust when bidders make small mistakes*, preprint, Stanford, 1999.

[7] J. Beirlant, J. Teugels, and P. Vynckier, *Practical analysis of extreme values*, K.U.Lueven, 2000.

[8] P. Bertail, C. Haefke, D. Politis, and H. White, *A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks*, UCSD, 2000.

[9] M. Buchinsky, *Changes in U.S. wage structure 1963-87: An application of quantile regression*, Econometrica **62** (1994), 405–458.

[10] R. Caballero and E. Engel, *Explaining investment dynamics in U.S. manufacturing: a generalized (S,s) approach*, Econometrica **67** (1999), no. 4, 783–826.

[11] G. Chamberlain, *Quantile regression censoring and the structure of wages*, Advances in Econometrics (C. Sims, ed.), Elsevier: New York, 1994.

[12] V. Chernozhukov, *Nonparametric extreme regression quantiles*, 1998, preprint, Stanford., 1998.

[13] ———, *Conditional extremes and near-extremes: Estimation, inference, and economic applications*, Ph.D. Dissertation, Stanford, 1999.

[14] V. Chernozhukov and H. Hong, *Robust feasible estimation in boundary-dependent models*, in progress, 2001.

[15] V. Chernozhukov and L. Umantsev, *Conditional value-at-risk: Aspects of modeling and estimation*, Empirical Economic **26** (2001), no. 1, 271–293, download http://www.mit.edu/ vchern/VaR.ps.

[16] B.J. Christensen and A. N. Kiefer, *The Exact Likelihood Function for an Empirical Job Search Model*, Econometric Theory **7** (1991), 464–486.

[17] J. Danielsson and C. G. de Vries, *Robust Tail Index and Quantile Estimation*, preprint, 1995.

[18] J. Davidson, *Stochastic limit theory*, The Clarendon Press, Oxford University Press, New York, 1994.

[19] L. de Haan, *Slow variation and characterization of domains of attraction*, Statistical extremes and applications (Vimeiro, 1983), Reidel, Dordrecht, 1984, pp. 31–48.

[20] A. Deaton, *The analysis of household surveys*, John Hopkins, Baltimore, 1997.

[21] A. Dekkers and L. de Haan, *On the estimation of the extreme-value index and large quantile estimation*, Ann. Statist. **17** (1989), no. 4, 1795–1832.

[22] S. Dolecki, G. Salinetti, and R. J. B. Wets, *Convergence of functions: equisemicontinuity*, Transactions of the Americal Mathematical Society **276** (1983), 409–429.

[23] S. G. Donald and H. J. Paarsch, *Maximum likelihood estimation when the support of the distribution depends upon some or all of the unknown parameters*, preprint. London, Canada: Department of Economics, University of Western Ontario, 1993.

[24] P. Doukhan, *Mixing*, Springer-Verlag, New York, 1994, Properties and examples.

[25] D. Duffie and J. Pan, *An Overview of Value at Risk*, preprint, 1997.

[26] P. Embrechts, C. Klüppelberg, and Milkosh. T., *Modelling extremal events*, Applications of Mathematics. Stochastic Modelling and Applied Probability, Springer, 1997.

[27] R. F. Engle and S. Manganelli, *Caviar: Conditional autoregressive value at risk by regression quantiles*, UCSD Economics Deaprtment Working Paper 99-20 , October, 1999.

[28] E. F. Fama, *The behavior of stock market prices*, Journal of Business **38** (1965), 34–105.

[29] P. D. Feigin and S. I. Resnick, *Limit distributions for linear programming time series estimators*, Stochastic Process. Appl. **51** (1994), no. 1, 135–165.

[30] B. Fitzenberger, *Computational aspects of censored quantile regression*, in L1-Statistical Procedures and Related Topics , ed. by Y. Dodge, IMS Lecture Notes-Monograph Series.

[31] C. Flinn and J. Heckman, *New methods for analyzing structural models of labor force dynamics*, Journal of Econometrics **18** (1982), 115–190.

[32] A.R. Gallant and H. White, *Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York: Bassil Blackwell, 1988.

[33] J. Galombos, *The asymptotic theory of extreme order statistics*, Wiley, Ney York, 1978.

[34] ______, *Order statistics. in:handbook of statistics. vol4, eds. p.r.krishnaiah and p.k.sen, pp. 359-382*, Amsterdam:North-Holland, 1984.

[35] C. Gourieroux, J.P. Laurent, and O. Scaillet, *Sensitivity analysis of values at risk*, Catholique de Louvain - I.R.E.S., 2000.

[36] E. Guerre, I. Perrigne, and Q. Vuong, *Optimal nonparametric estimation of first-price auctions*, Econometrica **68** (2000), no. 3, 525–574.

[37] G. Hall and J. Rust, *An empirical model of inventory investment by durable commodity intermediaries*, preprint, 1999.

[38] ______, *The (S, s) rule is an optimal trading strategy in a class of commodity price speculation problems*, preprint, 2000.

[39] P. Hall and C. C. Heyde, *Martingale limit theory and its application*, Academic Press, New York, 1980.

[40] B. M. Hill, *A simple general approach to inference about the tail of a distribution*, Ann. Statist. **3** (1975), no. 5, 1163–1174.

[41] H. Hong, *Non-regular maximum likelihood estimation in auction, job search and production frontier models*, mimeo., 1999.

[42] B. E. Honoré, *Simple estimation of a duration model with unobserved heterogeneity*, Econometrica **58** (1990), no. 2, 453–473.

[43] J. L. Horowitz, *Bootstrap methods for median regression models*, Department of Economics, University of Iowa, preprint, August 1996.

[44] O. Kallenberg, *Random measures*, fourth ed., Akademie-Verlag, Berlin, 1986.

[45] P. Kernes and A. Pagan, *Estimating the density tail index for financial time series*, Review of Economics and Statistics **79** (1997), no. 2, 171–175.

[46] K. Knight, *Epi-convergence and stochastic equisemicontinuity*, Preprint, 2000.

[47] ______, *Limit distributions of the linear programming estimator*, Preprint, 2000.

[48] R. Koenker and G. S. Bassett, *Regression quantiles*, Econometrica **46** (1978), 33–50.

[49] R. Koenker and O. Gelling, *Reappraising medfly longevity: A quantile regression survival analysis*, J. Amer. Statist. Assoc. (2001), to appear.

[50] R. Koenker and K. Hallock, *Quantile regression: An introduction*, preprint, download http://www.econ.uiuc.edu/research/, 2001.

[51] Roger Koenker and José A. F. Machado, *Goodness of fit and related inference processes for quantile regression*, J. Amer. Statist. Assoc. **94** (1999), no. 448, 1296–1310.

[52] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York-Berlin, 1983.

[53] R. M. Loynes, *Extreme values in uniformly mixing stationary stochastic processes*, Ann. Math. Statist. **36** (1965), 993–999.

[54] M. Mandelbrot, *Variation of certain speculative prices*, Journal of Business **36** (1963), 394–419.

[55] R. M. Meyer, *A poisson-type limit theorem for mixing sequences of dependent 'rare' events*, Ann. Probability **1** (1973).

[56] Christoffersen P. and Diebold F., *How relevant is volatility forecasting for financial risk management?*, Review of Economics and Statistics **82** (2000), 1–11.

[57] J. Pickands, *Statistical inference using extreme order statistics*, Ann. Statist. **3** (1975), 119–131.

[58] D. Politis, J. Romano, and M. Wolf, *Lecture notes on bootstrap and subsampling*, Stats Dept, Stanford (1999).

[59] S. Portnoy, *Using regression fractiles to identify outliers*, Y. Dodge (Ed) Statistical Data Analysis Based on L_1 Norm and Related Methods **North Holland, Amsterdam** (1987).

[60] S. Portnoy and J. Jureckova, *On extreme regression quantiles*, forthcoming in *Extremes*, 2001.

[61] J. Poterba and K. Rueben, *The distribution of public sector wage premia: New evidence using quantile regression methods*, NBER Working Paper 4734, 1995.

[62] J. L. Powell, *Censored regression quantiles*, Journal of Econometrics **32** (1986), 143–155.

[63] D. H. Pyle and S. J. Turnovsky, *Safety-first and expected utility maximization in mean-standard deviation portfolio analysis*, Review of Economics and Statistics **52** (1970), 75–81.

[64] R.D. Reiss, *Approximate distributions of order statistics*, Springer, 1989.

[65] S.I. Resnick, *Extreme values, regular variation, and point processes*, Springer, 1987.

[66] P. M. Robinson, *Nonparametric estimators for time series*, J. Time Ser. Anal. **4** (1983), no. 3, 185–207.

[67] R. T. Rockafellar and R.J. B. Wets, *Variational analysis*, Springer-Verlag, Berlin, 1998.

[68] A.D. Roy, *Safety first and the holding of assets*, Econometrica **20** (1952), 431–449.

[69] A. Sakov and P. Bickel, *An edgeworth expansion for the m out of n bootstrapped median.*, preprint, Berkeley, 1998.

[70] G. Salinetti and R.J. B. Wets, *On the convergence in distribution of measuarble multifunctions (random sets), normal integrands, stochastic processes and stochastic infima*, Mathematics of Operations Research **11** (1986), 385–419.

[71] H. Scarf, *The optimality of (S, s) policies in the dynamic inventory problem*, Mathematical methods in the social sciences, 1959, Stanford Univ. Press, Stanford, Calif., 1960, pp. 196–202.

[72] R. L. Smith, *Nonregular regression*, Biometrika **81** (1994), 173–183.

[73] R.L. Smith, *Extreme value analysis of environmental times series: an application to trend detection in ground-level ozone*, Statistical Science **4** (1989), 367–393.

[74] L. G. Telser, *Safety first and hedging*, The Review of Economic Studies **23** (1956), 1–16.

[75] R. Tsay, *Extreme value analysis of financial data*, preprint, 2000.

[76] A. W. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes*, Springer-Verlag, New York, 1996.

[77] V. Watts, H. Rootzén, and M. R. Leadbetter, *On limiting distributions of intermediate order statistics from stationary sequences*, Ann. Probab. **10** (1982), no. 3, 653–662.