# CONDITIONAL EXTREMES AND NEAR-EXTREMES: CONCEPTS, ASYMPTOTIC THEORY, AND ECONOMIC APPLICATIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ECONOMICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Victor V. Chernozhukov
August 2000

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Takeshi Amemiya
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thomas MaCurdy

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Patrick Bajari

Approved for the University Committee on Graduate Studies:

# Abstract

This dissertation develops the theory of extremal (near-extreme and extreme) quantile regression: estimation, inference, and applications in economics. In particular, the limit laws for extremal quantile regression statistics are obtained under conditions reflecting the inherently rare nature of the conditional extremal events, thus supplying a regression analogue of the classical extreme value theory. The relevant inference framework is provided. The results apply to cross-section as well as time series data. Examples, ranging from decision making under extreme uncertainty to price search models, provide motivation for and empirical applications of the theory developed.

# Acknowledgments

Best words can not express my gratitude to Takeshi Amemiya for teaching me Econometrics and for all of his devotion and relentless support of this dissertation. Permanent criticisms of Thomas MaCurdy helped me form practical views on Econometrics and the present work. I benefited greatly working as his research and teaching assistant. The always memorable midnight coffee sessions inspired my thinking about Econometrics and life in general. I am especially grateful to Mordecai Kurz for his support and wisdom. I am very indebted to Roger Koenker and Samuil Haimovich Aranson, Anil Bera, John Marden, Steven Portnoy. They made a pivotal contribution to my education and introduced Econometrics and Statistics to me. Conversations with Patrick Bajari helped me to motivate this research much better. I have learned considerably from beautiful papers of and conversations with Keith Knight, Jana Jurečkova, and Joseph Romano.

I am indebted to David Bullock, Lowell Hill, David Lins, Raymond Leuthold, Andrey Novotorov, Wayne Shafer, Bruce Sherrick, and Nicholas Yannelis for all of their attention that either helped me come to the U.S. or enter Stanford, and I am fortunate to have wonderful friends and co-workers: Han Hong, Robert McMillan, Nikolai Lehtinen, Sergei Morozov, Ljuda and Michael Parievsky, Lidiya Polukhina, Slavi Slavov, Tanya and Nikolai Tsepalov, and Len Umantsev. All-year-round mountain biking with Nikolai helped me think about extremes.

However, I am most grateful to my wife Ol'ga, my parents, and my brother for helping me to be myself. To them this dissertation is dedicated.

# Contents

**Abstract** **iv**

**Acknowledgments** **v**

**1 Introduction and Economic Motivation** **1**

1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1.2 Economic Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.3 The Concept of Extremes and Near-Extremes . . . . . . . . . . . . . . . . 7

1.3.1 What we wish to estimate and why . . . . . . . . . . . . . . . . . . 7

1.3.2 Linear Conditional Quantile Model . . . . . . . . . . . . . . . . . . 8

1.3.3 Sample Regression Order Statistics . . . . . . . . . . . . . . . . . . 8

1.3.4 Extremes, Near-Extremes & Data Scarcity . . . . . . . . . . . . . . 9

1.3.5 Basic Background on the Classical Extremes and Tail Types . . . . 11

1.3.6 Basic Assumptions of Our Analysis . . . . . . . . . . . . . . . . . . 14

**2 Asymptotics of Extreme Ranks** **16**

2.1 Essential Asymptotics . . . . . . . . . . . . . . . . . . . . . . . . . . 16

2.2 Generalization to Congenial Tail Heterogeneity . . . . . . . . . . . . . . 21

2.3 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

**3 Asymptotics of Intermediate Ranks** **24**

3.1 Essential Asymptotics . . . . . . . . . . . . . . . . . . . . . . . . . . 24

3.2 Generalizations to Congenial Tail Heterogeneity . . . . . . . . . . . . . . 27

3.3 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

**4 Extremal Processes and Related Inference** **29**
4.1 Introduction . . . . . 29
4.2 Definitions . . . . . 30
4.3 Behavior of Extremal RQ Processes with Extreme Ranks . . . . . 31
4.4 Behavior of Extremal RQ Process with Intermediate Rank . . . . . 32
4.5 Inference Process . . . . . 33
4.5.1 Inference Processes under Extreme Ranks . . . . . 33
4.5.2 Inference Processes under Intermediate ranks . . . . . 35
4.6 Conclusion . . . . . 37

**5 Inference** **38**
5.1 Introduction . . . . . 38
5.2 Inference about Tails . . . . . 38
5.2.1 The Regression Quantile Tail Index Estimator . . . . . 39
5.2.2 Estimation of the Tail Heterogeneity Index . . . . . 43
5.3 Confidence Intervals . . . . . 43
5.3.1 Inference under Extreme Ranks . . . . . 43
5.3.2 Inference under Intermediate Ranks . . . . . 46
5.4 Monte-Carlo Work . . . . . 46
5.4.1 Setup and Details . . . . . 47
5.4.2 Main Conclusion of Monte-Carlo . . . . . 49
5.5 Conclusion . . . . . 50

**Conclusion** **51**

**A Appendix for Chapters 2 and 3** **52**
A.1 Background Definitions and Lemmas . . . . . 52
A.1.1 Weak Convergence of Point Processes . . . . . 52
A.1.2 Convex Lower-Semi-Continuous Objectives . . . . . 54
A.2 Asymptotics under Extreme Ranks . . . . . 55
A.2.1 Preliminary Analysis and Basic Assumptions . . . . . 55
A.2.2 Asymptotic Distributions under Extreme Ranks . . . . . 58
A.2.3 Joint Convergence in Distribution . . . . . 62

A.2.4 Types of Limit Point Processes **N** under Heterogeneity and Dependence . . . . 62
A.3 Asymptotics under Intermediate Ranks . . . . 66
A.3.1 Asymptotic Distributions under Intermediate Ranks . . . . 66
A.3.2 Limits Under Tail Heterogeneity (and Serial Dependence) . . . . 67
A.4 Model of Congenial Tail Heterogeneity: Properties . . . . 70
A.5 Design Conditions for Extreme Ranks . . . . 74

**B Appendix for Chapter 4** **77**
B.1 Extreme Ranks . . . . 77
B.1.1 Useful Lemmas . . . . 77
B.1.2 Asymptotics for Extreme Ranks . . . . 78
B.2 Intermediate Ranks . . . . 80
B.2.1 Asymptotics under Intermediate Ranks . . . . 81
B.2.2 Sufficient Conditions for Stochastic Equicontinuity . . . . 84

**C Appendix for Chapter 5** **87**
C.1 Proofs . . . . 87
C.2 Monte-Carlo Illustrations . . . . 88
C.3 Estimating Unknown Variance: Comparison of Methods . . . . 99

**Bibliography** **101**

# List of Figures


1.1 Determinants of S,s-type adjustment rules, $f$ – conditional densities of the key state variable, $X$ – economic conditions. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.2 Simulation Example: True Densities, Extreme and Central Rank Approximations. QQ-plot of Extreme and Central Rank Approximations. On the density plot or QQ-plot, solid line denotes the true density or QQ-plot; line with triangles denotes the extreme rank approximation (almost indistinguishable from the true distribution); line with circles denotes central rank approximation. On the QQ-plot, the dashed line is the central approximation, and the dotted line is the extreme rank approximation. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

5.1 An Example of Tail Index estimation. The depicted is the estimator (function) $\hat{r}_\tau(m, l, \bar{X})$ for different values of $m$ and $l$ and $\tau$. On every plot '—' line corresponds to setting constant $m = 2$, '- - -' – $m = 1.5$ and '— — —' – $m = 1.2$. Plots correspond (from left to right, then down) to values of $l$ equal to 1.5, .5, 1.25, .75, 1.15, .9. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

C.1 Simulation Design: Marginal Densities of Regressors. . . . . . . . . . . . . . . . 92

C.2 Small Sample Approximation Properties. Case when $k = 1$, $T = 50$ . . . . . . . 93

C.3 Large Sample Approximation Properties. Case when $k = 1$, $T = 1000$. . . . . . . 94

C.4 Small Sample Approximation Properties. Case when $k = 5$, $T = 50$. . . . . . . . 95

C.5 Large Sample Approximation Properties. Case when $k = 5$, $T = 1000$. . . . . . . 96

C.6 Small Sample Approximation Properties. Case when $k = 15$, $T = 100$. . . . . . . 97

C.7 Large Sample Approximation Properties. Case when $k = 40$, $T = 1000$. . . . . . 98

# Chapter 1

# Introduction and Economic Motivation

## 1.1 Introduction

Models of conditional quantiles, invigorated by Koenker and Bassett[50], represent a substantially more general and informative method of regression analysis than conventional mean regression, since the former fully describe the conditional distribution of the response variable $Y$ given covariate $X$, without imposing any rigid distributional assumptions. The goal of this paper is to address the general issue of modeling *conditional or regression extremes,* namely the near-extreme and extreme regression quantile functions. At heart, conditional extremes and near-extremes model the extremal values of $Y$ conditional upon $X$. For example, the extreme (0th) conditional quantile function describes the minimum possible value of $Y$ given $X$ - that is, the conditional support function; the near-extreme 0.02 conditional quantile function describes the value below which $Y$ falls with probability 0.02 given values of $X$.

Modeling conditional extremes and near-extremes is essential for a diverse array of economic models. Some application include: (S,s) adjustment models of investment, technology update, sticky pricing, and cash-balance problems at the individual or aggregate level; reservation wage (price) strategies, commonly arising in dynamic job (price) search models; decision making under uncertainty, where the development of good market risk measures is vital for the purposes of insurance, management of risks, safety-first asset allocation and portfolio selection, financial intermediation, regulation, and resource management; production analysis, auction models and many others.

It should be noted that modeling non-regression extremal quantiles has been a subject of intense research for a number of years. See [56],[36],[37],[31] for important theoretical contributions that further encouraged exciting empirical observations. In contrast, conditional or regression extremal quantile modeling quantile has received substantially less consideration. Several recent papers by Smith[81], Jureckova and Portnoy[68], Knight[48],[49], Chernozhukov[17] should be mentioned. These papers studied asymptotics of sample extreme regression quantiles. Extreme regression quantiles shall be viewed here as a *very* special case of a more general paradigm: the *near-extreme and extreme* (or, as jointly called here, *extremal*) *regression quantiles*. In the no-covariates case, sample extreme quantiles correspond to the minimum (or maximum) order statistics. Hence they neglect the whole spectrum of statistics that lie in between – the $k$-th order (or rank) statistics. The sample near-extreme regression quantiles are seen here as precisely the $k$-th order *regression statistics*. Furthermore, any $k$-th order sample regression quantile (where we define order or rank to be $k = \tau T$, $\tau$ being the quantile index) can be considered as an "extreme" or "intermediate" or "central"(regular) order statistic, depending on how we treat its asymptotic behavior. In particular, the extreme and intermediate cases are defined as two forms of data scarcity: (i) intermediate order, when order is increasing, but at a slower rate than the sample size, and (ii) extreme order, when order is fixed relative to the data size. Note that in the regular case, the order is assumed to be growing at the sample size rate. The concepts are motivated through simple examples and have been previously used in analyzing the asymptotic behavior of order statistics in the non-regression problems by Von Mises, Fisher, Gnedenko, Smirnov, de Haan and others (see, for instance, [56], [31]). Importantly, tail inference in the no-covariate theory is also based on such concepts (see Pickands[61] and Hill[40]).

It must be stressed that the concepts of data scarcity (extremality) are applicable to the problem of conditional quantile estimation whenever data is finite (yet large) and the index of the estimated quantile is low or high in relation to the size of the data, as formalized above. Additionally, the problem of data scarcity is *amplified* by the presence of covariates. To that end, the concept of effective rank is useful. Effective rank, $r$, is the ratio of rank to the number of regressors, $k/d$. To motivate such a notion, consider the "regression" quantile problem in a sample of 1000 observations and 10 dummy regressors,[1] in which the target is the .1 conditional quantile function. The constructed estimate of slope will be the 10-th lowest order statistics in each of 10 subsamples corresponding to the dummy variables. The sample size is large, hence some asymptotic theory applies, but the question is *what* asymptotic theory? The above sources gave the answer for the non-regression problem. Our focus is to extend the theory to the regression settings.

Additionally, in the unconditional extreme value theory, the intermediate rank statistics play an immense inference role, providing the tail index estimators and the extrapolation quantile estimators ([61], [40], [26]). In our settings an analogous role is taken by the intermediate rank sample regression quantiles. This provides additional very powerful motivation for the theory we are developing.

In the present work, we systematically study the extreme and near-extreme (extremal) *regression* quantile statistics. To facilitate their use in the empirical work, we develop the asymptotic theory of extremal regression quantiles. We establish the asymptotic distributions and rates of convergence. Results allow for possibly dependent and heterogeneous data. Notwithstanding the results are novel for the simplest i.i.d. case. Congenial (from the inference prospective) models of tail heterogeneity and serial dependence are given. An inference framework is is also provided to provide completeness and immediate applicability.

[1]The setup is very close in nature to many studies in labor economics that have used quantile regression.

## 1.2 Economic Motivation

This work derives its empirical motivation from a variety of sources in economic research. Diverse theoretical and practical problems pose questions related to the extremal conditional quantiles, as illustrated in the following examples.

**Example 1.2.1 (Determinants of $(S, s)$ adjustment policies.)** In the basic form of $(S, s)$ adjustment models, an individual or a firm allows a state variable (the capital stock, for example) to fall until it reaches a certain *barrier* or *boundary*, denoted by $s$, at which point the stock is replenished to its upper value, $S$ (a boundary also). In many reasonable dynamic microeconomic scenarios, these policies are optimal (see Arrow et al[4] for a classical work on the subject, and, for instance, Caballero et al. [14],[13],[10],[11],[12]). The models are generalized by allowing the rules $(S, s)$ to depend on certain state variables $x$, such as firm characteristics, in which case these may be denoted as $(S(x), s(x))$. The importance of such a generalization is discussed in, e.g. [15]. Figure 1.1 illustrates the lower boundary $s(x)$ of this kind. A recent model by Caballero and Engel [11] allows for "stochastic bands," in which case the upper and lower near-extreme conditional quantile functions describe the approximate adjustment "rules", or hazard functions.[2] Figure 1.1 illustrates such a situation for the upper band: .98 conditional quantile function signifies the level at which in 98% of all 'cases' the key stock variable is already adjusted.

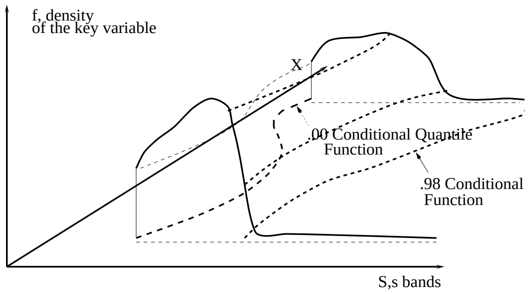


Figure 1.1: Determinants of S,s-type adjustment rules, $f$ – conditional densities of the key state variable, $X$ – economic conditions.

Some specialized examples associated with such models include: capital stock adjustment,

[2] In the empirical settings, this model, may be preferred to the pure $(S, s)$ model on robustness considerations .

technology update, sticky pricing models, cash-balance, and others problems.

**Example 1.2.2 (Decision Making in the Presence of Extreme Uncertainty)** Study of risk is a key subject of many investment decisions, insurance, safety-first asset allocation and portfolio selection, financial intermediation, regulation, and resource management. The approach, called *value-at-risk* , seeks to model high level risk that occurs infrequently.[3]

Two basic issues in the analysis of risk deserve special mention. Market participants, on the one hand, allocate funds among risky investment categories (equity, bonds, foreign exchange, derivative securities), and are specifically concerned about large extremely negative returns that can wipe out their capital and cause illiquidity or insolvency. On the other hand, the regulatory institutions may set limits on trade to prevent extreme movements in asset prices that can threaten liquidity of financial intermediaries and generate widespread panic. In so doing, government agencies must rely on probabilistic evaluation of such rare events within either static or dynamic market environment. In the dynamic setting, a *conditional* distribution (its extremal part) has to be used, as opposed to the unconditional static settings. Indeed, it is a common *practice* for both the internal and external management and for the regulation purposes to compute value-at-risk measures that reflect the threshold levels below which the return is only 1% or 5% of the time. To translate this practice into the context of *conditional* risk assessment is one purpose of our research.

An important branch of the literature that justifies the analysis of conditional extreme events is the seminal work of Roy[75], Telser[82], Pyle and Turnovsky[70], and others, addressing *safety-first* portfolio selection. In this approach, the investors solve one of the following problems:

1. $\min Pr(Z \leq \bar{z}|I_t)$,
2. $\max\limits_{Pr(Z\leq\bar{z}|I_t)\leq\alpha} \bar{z}$,
3. $\max\limits_{Pr(Z\leq\bar{z}|I_t)\leq\alpha} \mu$,

where $Z$ is a random return on the portfolio, $\bar{z}$ is the subsistence or disaster level of return, $\alpha$ is the probability of a disaster, set to be small, $\mu$ - mean return of the portfolio, and $I_t$

[3]See, for instance, Lo and Ait-Sahalia[1], and also [20], [83]. Duffie and Pan[30] offer a survey of V-a-R modeling techniques.

is the conditioning vector of information variables at a given date $t$. In all cases, for small values of $\alpha$, liquidity preference behavior of investors is strongly linked to the conditional distribution of extremal events – conditional extremal quantiles. Conceivably, safety-first analysis also includes transaction demand as well as the investment demand for assets.

The tail shape of payoff distribution (equivalently extremal quantile shape) has been found to be important in a number of other contexts as well. In particular, Mandelbrot[58], Fama[32], Kearns and Pagan[47], Danielsson and de Vries[23], to name a few, study the shape of distribution of prices and returns and stress the important nature of tail behavior, or extremal quantile for decision making. [4]

**Example 1.2.3 (Reservation Prices in Dynamic Search Models)** Many models of job search predict a reservation wage policy as the optimal strategy of the worker – wages are accepted if they exceed certain threshold levels (conditional minimum wages), which are functions $W$ of $X$. Hence the observed accepted wage distribution is the original offer wage distribution truncated from below at $W(X)$. See e.g. Sargent[77]. Heckman and MaCurdy [39] offer an excellent survey of labor economics models that result in reservation wage policies.[5] Other interesting models where reservation wages (prices) play an important role may be found in Christensen and Kiefer[21], Flinn and Heckman[34] and Weitzman [88].

Modeling $W(X)$ as merely a boundary, i.e. an extreme conditional quantile function, is probably not desirable due to the presence of outliers and misreporting. However, it should be noted that the (approximate) reservation wage can be well described in terms of a conditional quantile function - 0.01 or 0.02 conditional quantiles describe the wage threshold levels for 99 or 98 percent of the labor force with the given socioeconomic status and given economic environment $X$. Therefore, estimation and inference about the approximate reservation wages as well conditional poverty levels may be accomplished in terms of near-extreme quantile regressions.

**Example 1.2.4 (Auctions)** Consider a first-price procurement auction where each of $n$ risk neutral bidders has a (privately observed) cost of completing the project, $c$, independently and identically distributed according to the law $G$. A symmetric pure strategy

[4] The tails of the payoff (utility) distribution plays an important role in the recent robust and risk-sensitive decision making theory, where the optimal behavior of the agent is derived through the max-min expected utility framework. See e.g.[3] and references therein.

[5] Some notable applications of quantile regression in wage analysis include Chamberlain[16], Buchinsky[9], Poterba and Rueben[69], and others.

Bayesian Nash equilibrium bidding rule, $\sigma(c)$, satisfies:

$$\sigma(c) = c + \frac{\int_c^\infty (1 - G(\xi))^n d\xi}{(1 - G(c))^n}.$$

Assuming that $G(c)$ belongs to a parametric family $G(C, \theta)$, the lower boundary of the observed bid distribution, which is the bid made by the buyer with the lowest possible cost $c = 0$, is $\int_0^\infty (1 - G(\xi, \theta))^n d\xi \equiv \mathcal{E}(n, \theta)$. The lower boundary of the distribution of the observed bid data - the object of econometric interest - depends on both the distribution parameters and the number of buyers. This frontier, therefore, is *environment-dependent.* The truncation problem arising in auction models, similar to the setup we described, has been analyzed in remarkable works of Donald and Paarsch[28] and Hong[43], Hong and Shum[44]. Both in [28] and [43], maximum likelihood estimation procedures explicitly incorporate the conditional support modeling. Feasible ML estimators that incorporate the extreme and near-extreme regression quantile estimates of the boundary functions could be considered by substituting them in the likelihood function. Just as the direct ML approach, the feasible estimation with the extreme regression quantile estimates[6] may suffers from acute unrobustness (see Bajari[5].) In such instances, the problem could often be mitigated by considering "feasible" estimation with the near-extreme regression quantiles that do not fit the outliers. The resulting asymptotics of the MLE will be qualitatively similar due to the nature of asymptotics of the near-extremes that we derive here.[7][8]

**Example 1.2.5 (Other Models)** The applications of the extremal conditional quantiles are not limited to those listed above. *Any* economic model that strives to explain extremal contingencies conditional on the observed covariates is amenable to our tools.[9]

## 1.3 The Concept of Extremes and Near-Extremes

### 1.3.1 What we wish to estimate and why

We work with a sequence of random variables $\{Y_t, X_t\}$ taking values in $\mathbb{R} \times \mathbb{R}^d$. $Y_t$ is the response variable, $X_t$ is the conditioning variable, the covariates.

[6] It can be argued that in many situations the two will be asymptotically equivalent – e.g. [81], [43] .

[7] Under the extreme rank condition – see next section.

[8] Consequently, the results on feasible estimation e.g. in Smith[81] remain valid even when the robust near-extreme quantile regressions are used as boundary estimates. By operation "substituting", we mean that all the observations below and on the estimated boundary function are deleted from the likelihood.

[9] We refer the reader to [31], [36], [37], [56] for many more examples in other areas of science.

The $\tau$-th conditional quantile function $Q_{Y_t}(\tau|X_t)$ is a function $m(X_t)$ that satisfies the relationship $P(Y_t \leq m(X_t)|X_t) = \tau$. For instance, $Q_{Y_t}(.5|X_t)$ is the conditional median, $Q_{Y_t}(.1|X_t)$ is the conditional decile, and $Q_{Y_t}(0|X_t)$ is the conditional extreme (minimal) regression quantile or the conditional lower support function. If $F_{Y_t}(\cdot|x)$, the distribution function of $Y_t$ conditional on $X_t = x$, is continuous and strictly monotone, $Q_{Y_t}(\cdot|X_t)$ is its inverse and hence unique. More generally,

$$Q_{Y_t}(\tau|X_t) \equiv F_{Y_t}^{-1}(\tau|X_t),$$

where $F_{Y_t}^{-1}(\cdot|X_t)$ is the (generalized[10]) inverse of $F_{Y_t}(\cdot|X_t)$. Our focus is exclusively on modeling[11]

$$F_{Y_t}^{-1}(\tau|X_t), \text{ where } \tau \text{ is } 0 \text{ or } \textit{close} \text{ to } 0,$$

that is, the *extreme* and *near-extreme* (hereafter **extremal**) conditional quantile functions. The notion of *closeness* will be build in relation to the sample size.

### 1.3.2 Linear Conditional Quantile Model

We consider a linear (polynomial) model:

$$Q_{Y_t}(\tau|X_t) = F_{Y_t}^{-1}(\tau|X_t) = X_t'\beta(\tau), \quad \forall \tau. \tag{1.3.1}$$

This model is implied, for instance, by the linear heteroscedastic location-scale regression models, and is very flexible in assuming little about the underlying distribution. Note that $X_t$ may incorporate a wide array of polynomial and alternative transformations of the observable covariates. Furthermore, since the conditional quantile function is the inverse of the conditional distribution function, it is a *complete summary* of regression relationship.

### 1.3.3 Sample Regression Order Statistics

Suppose we have $T$ observations $\{Y_t, X_t\}$. In the no-covariates case, the $\lceil \tau T \rceil$-th order statistic,[12] $\hat{\beta}(\tau)$, or sample $\tau$-th quantile, is generated by solving the problem:

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

[10] see [55]

[11] And cases when $\tau = 1$ or close to 1. W.l.o.g. we discuss only the case above.

[12] $\lceil \tau T \rceil$ denotes the least integer greater or equal to $\tau T$.

where $\rho_\tau(x) \equiv (\tau - 1(x \le 0))\, x$.

Koenker and Bassett[50] proposed to extend the concept to the regression setting by solving an analogous problem involving the covariates:[13]

$$\min_{\beta \in \mathbb{R}^d} \quad \sum_{t=1}^{T} \rho_\tau \left( Y_t - X_t' \beta \right). \tag{1.3.2}$$

### 1.3.4 Extremes, Near-Extremes & Data Scarcity

*Formal Concepts of Data Scarcity*

We view the sample regression quantiles as a natural generalization of order statistics in regression settings. Hence extremal regression quantiles are seen here as the low order conditional order statistics. To be exact, for any given sample size $T$, the $\tau$-th regression quantile fit are seen here as the $\lceil \tau T \rceil$-th *conditional* order statistic. Henceforth, we shall refer to $\tau T$ as to the *rank* or *order*. Throughout, without loss of generality, we consider indices $\tau < 1/2$.

We now *informally* distinguish two types of data scarcity or extremality:

(i) extreme order (rank) case, when $\tau$ is very low relative to $T$ (very far into the tail),

(ii) intermediate order (rank) case, when $\tau$ is low relative to $T$ (not very far into the tail).

In contrast, the regular case involves the central and close to central ranks (close to center of distribution). These notions make $\tau$ sample size dependent: $\tau = \tau_T$. However, to ease the notational burden, write $\tau$ instead of $\tau_T$. *Formally*, the following *asymptotic* conditions are employed to *capture* the notions of data scarcity:

(i) extreme order (rank) case: $\tau$ is s.t. $\tau \searrow 0$, $\tau T \to k > 0$,

(ii) intermediate order (rank) case: $\tau$ is s.t. $\tau \searrow 0$, $\tau T \to \infty$.

In non-regression settings, the behavior of sample quantiles under both the extreme and intermediate order cases has been extensively researched. It forms an essential part of the (non-regression) *extreme value theory*. The analysis of the extreme order case started in

[13] They gave many arguments as to why this gives a good regression analogue of the order statistics. Also note that former and the latter objective functions can be seen as integrations of the sample moment condition: $\frac{1}{T}\sum_{t=1}^{T} (\tau - 1(Y_t \le X_t'\beta))\, X_t \approx 0$, that should approximately hold at the true parameter value $\beta(\tau)$, since $E([\tau - 1(Y_t \le X_t'\beta(\tau))]X_t | X_t) = 0$.

the 1920s and still continues. This includes the remarkable results on the distribution of extremes, obtained by Fisher, Frechet, Tippett, Gnedenko, Chibisov, Smirnov, Von Mises, Galombos, de Haan and others (mainly between the 1930s and 1960s) – (see [31] for an excellent review.) Several prominent monographs focus on the extreme order statistics – Galombos[36], [37], Leadbetter, Lindgren, and Rootzen[56], and Resnick[71]. The intermediate order statistics in non-regression settings were considered in a celebrated paper by Chibisov[22] and Smirnov[79] that appeared in the 1960s. Recently, in the 1980-90s a series of papers continued to address the subject (see Dekkers and de Haan [26]).

The contribution of our work is to extend the previous achievements to the *regression setting*, that is more interesting in economics.

*A Motivating Example for the Principles (i) and (ii)*

There are three principal motivations for the data scarcity principles (i) and (ii):

1. The principles allow to obtain alternative, better approximations of sample extremal quantiles than the central rank theory, when nuisance parameters are known;

2. Given these alternatives, the principles provide ways to robustly approximate the true distributions when nuisance parameters are unknown. Nuisance parameters can be estimated better than by centrality based methods. In order to construct variance estimate, central rank or regular theory of inference requires the nonparametric estimate of the conditional density function evaluated at the near-extreme quantile. In contrast, intermediate rank theory requires only the estimate of tail index and a tail heterogeneity index.[14] Extreme rank theory enjoys similar congenial properties. As a result, the inference about extremal quantiles is much more robust when using the central theory[15].

3. The principles provide for an additional inferential role. Consistent tail index estimators, tail heterogeneity index estimators, extrapolation quantiles are all based on the extremal quantile (regression) statistics under the intermediate ranks (Chapter 5).

[14] Tail heterogeneity parameter is a new term introduced in this paper. It measures how strongly covariates affect the tail behavior. It is a product of assuming linear conditional quantile model and the extremal conditional tails (in the domain of minimum attraction). See Chapter 2.

[15] Appendix C supports this with a Monte-Carlo example.

It is useful to illustrate the first point. Suppose we have sample of i.i.d. random variables $U_t$, $t = 1, 100$, and wish to infer about the .03-th quantile of the underlying distribution. A natural estimator[16] is $U_{(6)}$, the sixth order statistic in the sample of size 200. Given that the sample is large, consider the asymptotic approximations to the distribution of this estimator. On the one hand, regular asymptotics (e.g. in Mosteller[60], Koenker and Bassett[50]) states:

$$\sqrt{T}(\widehat{F}_u^{-1}(.03) - F_u^{-1}(.03)) \xrightarrow{d} N(0, \sigma^2),$$

for some variance $\sigma^2$. On the other hand, the estimator is only the sixth order (*nearly-extreme*) statistic and (for example, taking $F_u$ with unbounded support and algebraic tails: $F_u(y) \sim y^{-1}$ as $y \to -\infty$):

$$\frac{1}{T}(U_{(6)} - F_u^{-1}(6/T)) \xrightarrow{d} K - (\mathcal{E}_1 + ... + \mathcal{E}_6)^{-1},$$
$$\mathcal{E}_i \text{ i.i.d. standard exponential, } K \text{ is some constant,}$$

as can be shown using results of classical extreme value theory.[17] That is, the limit distribution is an extreme value distribution. Therefore, the central rank approximation tells that $T^{\frac{1}{2}}(\widehat{F}_u^{-1}(.03) - F_u^{-1}(.03)) \overset{\mathcal{D}}{\approx} N(0, \sigma^2)$. The extreme rank approximation suggests, on the other hand ($T = 100$): $T^{-1}(\widehat{F}_u^{-1}(.03) - F_u^{-1}(.03)) \overset{\mathcal{D}}{\approx} K - (\mathcal{E}_1 + ... + \mathcal{E}_6)^{-1/2}$. Figure 1.2 *clearly favors* the second approximation.

Discussion of the other motivation points (quality of inference and the inferential role of near-extreme quantiles) is further discussed in Chapter 5, appendix C, and [18].

### 1.3.5 Basic Background on the Classical Extremes and Tail Types

We now sketch the definitions and basic results that are important for understanding the rest of the paper. In the subsequent analysis, we must make some assumptions about the tail behavior.[18]

[16] Natural may not mean the best. Extrapolation estimators defined on the basis of intermediate order sample regression quantiles may be preferred to the sample analogues. See section [18]. Here we focus on building the distribution theory for the conditional order statistics.

[17] See Theorems 1 and 2.

[18] Restrictions on tail shapes made by the classical theory are mild and they are justifiable from the view point of necessitating the (asymptotic) statistical decision making. The recent work by Schwarz[78], using invariance rationality conditions, derives restrictions on the tail behavior of priors in the Savage decision making framework under extreme uncertainty. Remarkably, this theory predicts the tails of one of the extremal types.

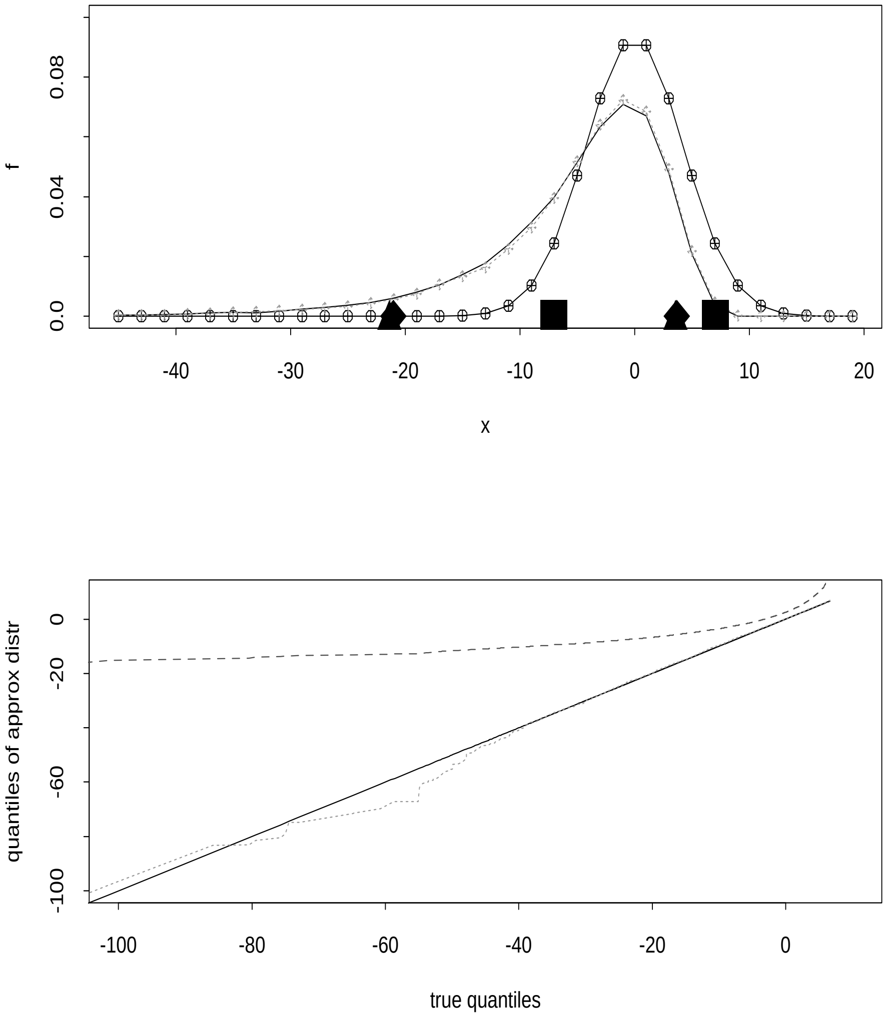


Figure 1.2: Simulation Example: True Densities, Extreme and Central Rank Approximations. QQ-plot of Extreme and Central Rank Approximations. On the density plot or QQ-plot, solid line denotes the true density or QQ-plot; line with triangles denotes the extreme rank approximation (almost indistinguishable from the true distribution); line with circles denotes central rank approximation. On the QQ-plot, the dashed line is the central approximation, and the dotted line is the extreme rank approximation.

*Types of Support*

Two basic forms of tails of $F_Y(\cdot|X_t)$ are those with:

- finite support (boundary or frontier case), $F_{Y_t}^{-1}(0|X_t) > -\infty$,
- infinite support $F_{Y_t}^{-1}(0|X_t) = -\infty$.

We treat *both* cases, and both cases can often be considered as approximations of each other.

*Extremal Tail Types and Extreme Rank Statistics*

Consider the order statistics $U_{(1)}, ..., U_{(k)}$ from the i.i.d. sample $U_1, ..., U_T$, distributed according to law $F_u$, with the lower end-point given by $x_{F_u} = 0$ or $x_{F_u} = -\infty$. The question posed by the classical non-regression extreme value theory was to describe existence and the forms of the non-degenerate limit laws for the properly normalized extreme order statistics:

$$a_T(U_{(k)} - b_T) \xrightarrow{d} \quad ?, \quad a_T = \ ?, \quad b_T = \ ? \ .$$

Following work of Frechet, Fisher and Tippett, Gnedenko, and later de Haan (see e.g. [56]), answered the question. At first, the focus is on $U_{(1)}$. The asymptotic laws of $U_{(1)}$ were identified as: [19]

$$H_\xi(x) = e^{-(1+x\xi)^{-1/\xi}}, \ 1 + \xi x > 0, \quad \xi \in \mathbb{R},$$

that for $\xi = 0, \xi > 0, \xi < 0$ define respectively *type 1, 2, and 3 extreme value distribution* functions (Gumbel, Frechet, and Weibull). The necessary and sufficient restriction on $F_u$ s.t. the limit law is $H_\xi$ is that $F_u$ has *extremal type 1-3 tails*:[20][21]

$$\begin{aligned}
&\text{type 1,}(\ \xi = 0\ ) &&: \text{ as } t \searrow 0 \text{ or } -\infty, && F_u(t + x\ell(t)) \sim F_u(t)e^{-x}, && \forall x \in \mathbb{R},\\
&\text{type 2, } (\xi = \frac{1}{\alpha}) &&: \text{ as } t \searrow -\infty, && F_u(t) \sim F_u(tx)x^{\alpha}, && \forall x > 0,\ \alpha > 0,\\
&\text{type 3, } (\xi = \frac{-1}{\alpha}) &&: \text{ as } t \searrow 0, && F_u(t) \sim F_u(tx)x^{-\alpha}, && \forall x > 0,\ \alpha > 0.
\end{aligned} \tag{1.3.3}$$

Type 1 class includes exponentially and near-exponentially tailed distributions; type 2 class – algebraically and near-algebraically tailed at $-\infty$ (regularly varying at $-\infty$); type 3

[19] All location-scale transformations $ax + b (a > 0)$ of the argument of the limit c.d.f.s are permitted.

[20] $A \sim B$ if $A/B \to 1$.

[21] Without loss, an appropriate choice of function $\ell$ for type 1 is given by $\ell(t) = \int_{x_{F_u}}^{t} F_u(v)dv/F_u(t)$, for $t > x_{F_u}$ (*cf.*[56]).

class – algebraically and near-algebraically tailed at a finite end point, taken as 0 (regularly varying at 0).[22]

The *canonical* normalization constants are given by:

$$\begin{array}{llll} \text{type 1:} & a_T \sim & 1/\ell\big[F_u^{-1}(\tfrac{1}{T})\big], & b_T = F_u^{-1}(\tfrac{1}{T}), \\ \text{type 2:} & a_T \sim & -1/F_u^{-1}(\tfrac{1}{T})\ , & b_T = 0, \\ \text{type 3:} & a_T \sim & 1/F_u^{-1}(\tfrac{1}{T})\ , & b_T = 0. \end{array} \tag{1.3.4}$$

The limit results for $k$-th order statistic, with fixed $k$, $U_{(k)}$, have been then identified in terms of $H_\xi$, with the same normalization constants (cf. [56]).

*Intermediate Rank Statistics and Extremal Tail Types*

As discussed, one of the most general and fairly recent treatments of the intermediate order statistics is the work of Dekkers and de Haan[26]. Using slightly stronger restrictions on the tails, discussed in below, they found that the limit laws are normal, but variance *depends on the extremal tail types* through the tail index $\xi$, as $k = [\tau T] \to \infty$[23] (Theorem 3.1) [24]

$$\left[\sqrt{\tau T}\frac{1}{F_u^{-1}(2\tau) - F_u^{-1}(\tau)}\right]\Big(U_{([\tau T])} - F_u^{-1}(\tau)\Big) \xrightarrow{d} N\left[\ 0,\ \frac{\xi^2}{(2^{-\xi}-1)^2}\ \right]. \tag{1.3.5}$$

### 1.3.6 Basic Assumptions of Our Analysis

We consider a probability space $(\Omega, \mathcal{F}, P)$, possibly indexed by $T$, and we assume that $X_t$, the properly *scaled* regressor,[25] includes a constant so that the no-covariate case is

[22] Note some examples. Type 1 class includes all exponentially tailed distributions, such as normal or any distribution that exhibits the exponential tails near the end-point (assumed 0 throughout), e.g. $F_u(x) = K_1 \exp[-K_2/(x - x_F)], x > x_F, K_1, K_2 > 0$. Type 2 class includes distributions with algebraic tail at $-\infty$, for instance, all $F_u$ s.t. $F_u(x) \sim K(-x)^{-\alpha}, \quad \text{as } x \searrow -\infty, \quad K > 0, \quad \alpha > 0$. Type 3 class includes all truncated distributions (e.g. truncated at 0 normal) and, for instance, those that behave algebraically near the end-pont 0: $F_u(x) \sim Kx^\alpha, \quad \text{as } x \searrow 0, \quad K > 0, \quad \alpha > 0$.

[23] Note that we have rephrased the result in terms of low and not high quantiles as stated in [26]

[24] In fact the scaling $a_T$ was found to be $\sqrt{\tau T}/(U_{(2[\tau T])} - U_{([\tau T])})$. This can be replaced by $\sqrt{\tau T}/(F_u^{-1}(2\tau) - F_u^{-1}(\tau))$, since the ratio of the two converges to 1 in probability. Pickands[61] has shown this as a proof of consistency of his estimator of tail index $\xi$.

[25] $X_t$ and the linear conditional quantile model are assumed to be linearly transformed so that (at the least) the limit empirical distribution function of $X_t$ is non-degenerate. If $X_t$ are i.i.d. nondegenerate random variables, no scaling may be necessary. However, for instance, if $X_{jt} = t$ (linear trend), redefine $X_{jt}$ as $t/T$ and the corresponding quantile coefficient $\beta(\tau)_j$ as $\beta(\tau)_j \times T$ (by linearity). This is possible also in finite sample due to equivariance of the canonical sample regression quantiles (see [50]). This rescaling allows for

incorporated as the special case. Define $U_t = Y_t - X_t'\beta(q)$, where $q = 1/2$ and $q = 0$ in the infinite and finite support cases respectively.

***Assumption 1***

*(A)* $F_{U_t}(z|X_t = x) \sim F_u(z)$[26], *as* $z \searrow -\infty$ *(infinite support) or as* $z \searrow 0$ *(finite support), uniformly in* $x \in \mathbf{X}$, $t$. $F_u$ *has type 1,2, or 3 tails.*

*(B) Support of (properly scaled)* $X_t$ *is in (or trimmed to) a compact subset* $\mathbf{X}$ *of* $\mathbb{R}^d$.[27]

*(C)* $\{Y_t, X_t\}$ *are independent across* $t$*, and*

*(D) Limit empirical distribution function of* $X_t$, $F_X$, *with support* $\mathbf{X}$ *and mean* $\mu_X$, *exists and is nondegenerate in* $\mathbb{R}^d$. *For at least one* $i$, $F_{X_i}$ *is absolutely continuous (if* $d \geq 2$*).*

Note that the key assumption A requires the tail of conditional distribution to be *asymptotically* independent of covariate. This is, no doubt, quite restrictive. However, this simplest set of conditions will be the focus of the discussion to *convey the essential results as ideas*, not some dry formalisms. In subsequent sections and appendix we will consider

- tail heterogeneity in the covariate,
- more general design,
- dependent data.

It should be stressed that our general treatment and derivations *do not rely* on a particular form of data dependence or heterogeneity assumption. In the appendix, we *require only* the weak convergence of a key point process (to be defined), and under intermediate ranks – certain LLN, CLT, and analytical properties. Using these general results, we provide the explicit limit theorems under sets of sufficient conditions that rather flexibly allow for these more general conditions. But the main results of the appendix allow extensions of these sufficient conditions to many other directions by supplying *only* the limit theorems on weak convergence of point processes, CLTs, and LLNs, as required.

various forms of trends and other non-stable regressors. We will not go back to this issue and always assume onwards that $X_t$ have already been transformed in this way.

[26] As a reminder, $A \sim B$ if $A/B \to 1$.

[27] That is, in case of unbounded support, we sample only those observations $Y_t, X_t$ s.t. $X_t \in \mathbf{X}$.

# Chapter 2

# Asymptotics of Extreme Ranks

## 2.1 Essential Asymptotics

The approximation concept here involves the case when $\tau T \to k > 0$. In what follows, we confine our discussion to type 3 tails and only briefly state the results for type 1 and 2 tails. Our intent here is to *sketch the main ideas*, so that the key structure is clear. A substantively more general treatment of the problem and technical matters is deferred to the appendix.

The scheme of the argument is to obtain a finite-dimensional weak limit $Q_\infty(\cdot)$ of the sequence of suitable objective functions $\{Q_T(\cdot)\}$. Then properly normalized regression quantile statistics, $\hat{Z}_T$, defined as an argmin of $Q_T(\cdot)$, will weakly converge to random variable $Z_\infty$, the argmin of $Q_\infty(\cdot)$, by virtue of nice properties of $\{Q_T(\cdot)\}$ and $Q_\infty(\cdot)$.

We form a convenient objective function, rephrasing also the earlier definitions in the form of the appropriately rescaled processes. Define

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

Then $\hat{Z}_T$ is the optimum of the rescaled objective function in (1.3.4):

$$Q_T(z) \equiv \sum_{t=1}^{T} \big(\tau - 1(U_t a_T \le X_t' z)\big)\big(U_t a_T - X_t' z\big) - \tau U_t a_T. \tag{2.1.1}$$

where $U_t \equiv Y_t - X_t'\beta(0) \ge 0$. Note that we have subtracted term $\sum_t \tau U_t a_T$, which is critical in constructing some nice properties, as to be seen.[1] Clearly this does not affect the argmin

[1]This formulation only concerns the type 3 case. Similar but not direct extension to the tail types 1 and 2 is stated at the end of this section, while the proofs appear in appendices.

$\hat{Z}_T$. That is,

$$Q_T(z) = -T\tau\bar{X}'z - \sum_{t=1}^{T} 1\big(U_t a_T \le X_t'z\big)\big(U_t a_T - X_t'z\big). \tag{2.1.2}$$

This objective function is convenient since it is continuous and convex. Furthermore, $Q_T(\cdot)$ is constructed as a convenient functional of the point process to be defined next. This allows us to derive the finite-dimensional (*fidi*) weak limit of $Q_T(\cdot)$ employing the point process methods.

The fidi distribution of $Q_T(\cdot)$ is defined by the distribution of its coordinates $\{Q_T(z_1), ..., Q_T(z_l)\}$ for any $z_1, ..., z_l$. Since $\bar{X} \xrightarrow{P} \mu_X$, and $\tau T \to k$ (for $j \le l$) :

$$Q_T(z_j) = -k\mu_X' z_j - \sum_{t=1}^{T} 1(U_t a_T \le X_t z_j)\big(U_t a_T - X_t' z_j\big) + o_P(1). \tag{2.1.3}$$

We will show that the limit behavior of the summation term in (2.1.3) is *determined* by the *point process* (*random measure*) $\widehat{\mathbf{N}}$ that assigns mass to a (measurable) set $A$ by:

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

Roughly speaking, a process $N$ is a point process, if it is a measure that counts the number of random points falling in a given measurable set $A$.[2] I.e. $N$ is a random counting measure with atoms at its random points (above, these are $(a_T U_t, X_t), t = 1, ..., T$ ). One defines Lebesgue-Stieltjes integral $\int g dN$ of process $N$ with points $\{X_i\}$ as:

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

The convergence of such integrals (where $g$ are usually taken to be positive and vanish outside a compact set) is the key to defining weak convergence of the point processes. Formalizations and treatments of weak convergence of random measures can be found e.g. in [46] (see also Appendix.)

It is now possible to represent $Q_T(z_j), j \le l$ as an integral w.r.t. $\widehat{\mathbf{N}}$:

$$Q_T(z_j) = -k\mu_X' z_j + \int (l - x'z_j)^- d\widehat{\mathbf{N}}(l, x) + o_P(1), \tag{2.1.4}$$

[2] See [46] or [72] for definitions

$\big(Q_T(z), j \leq l\big)$ is a continuous map of the point process $\widehat{\mathbf{N}}$. Hence its weak convergence is determined by the convergence of point process $\widehat{\mathbf{N}}$ by continuous mapping theorem. Let us note that verification of continuity, relegated to appendix, is a main part of proof. It relies on a *careful construction* of *point process* $\widehat{\mathbf{N}}$ and of *space* $E$ (and its *topology*) on which $\widehat{\mathbf{N}}$ is defined, for each of tail types.

The point process $\widehat{\mathbf{N}}$ converges weakly to a Poisson Point Process $\mathbf{N}$, defined by [3]

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

where $\{J_i\}$ are random points defined as:

$$\begin{aligned} J_i &\equiv \Gamma_i^{\frac{1}{\alpha}}, \qquad i \geq 1 \\ \Gamma_i &\equiv \mathcal{E}_1 + ... + \mathcal{E}_i, \end{aligned} \tag{2.1.5}$$

$\{\mathcal{E}_i\}$ are i.i.d. exponential random variables with mean 1, and $\{X_i\}$ are i.i.d. with law $F_X$, distributed independently of $\{\mathcal{E}_i\}$. A more general treatment in the appendix allows for a richer variety of limit processes.

Hence we conclude that the fidi weak limit of $\{Q_T(\cdot)\}$ 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 the convexity theorem 5 for semi-continuous objectives in Knight[48]) the limiting distribution for $\hat{Z}_T$, provided $Q_\infty(\cdot)$ has unique argmin a.s. and is finite on an open non-empty set. Careful treatment of these issues is in the appendix. Hence

$$a_T\Big(\hat{\beta}(\tau) - \beta(0)\Big) \xrightarrow{d} Z_\infty \equiv \operatorname*{argmin}_z Q_\infty(z), \text{ as } \tau T \to k, T \to \infty. \tag{2.1.6}$$

[3] For the reader not familiar with point process methods, it is perhaps instructive to see some intuition. Note that $\widehat{\mathbf{N}}$ can be rewritten as:

$$\widehat{\mathbf{N}}(A) = \sum_{i=1}^{T} 1\Big(\{\ a_T U_{(i)}, X_{\{i\}}\} \in A\ \Big)$$

where $U_{(i)}$ is $i$-th rank error and $X_{\{i\}}$ is the corresponding to it $i$-th covariate vector. Note that $a_T U_{(1)}$ is asymptotically distributed as $(\mathcal{E}_1)^{\frac{1}{\alpha}}$, ..., $a_T U_{(l)}$ – as $(\mathcal{E}_1 + ... + \mathcal{E}_l)^{\frac{1}{\alpha}}$ (these are standard results in extreme value theory). Since $a_T$ increases, $1(\{\ a_T U_{(i)}, X_{\{i\}}\} \in A)$ is non-zero only for a few low rank errors if $A$ is compact. Hence $\mathbf{N}$ may be considered a good approximation to distribution of $\widehat{\mathbf{N}}$ for large $T$. A formal proof is a special case of the tail-heterogeneous case in the appendix.

Further, it is shown that $a_T(\beta(0)-\beta(\tau)) \to k^{\frac{1}{\alpha}}\mathbf{e}_1$ for $\mathbf{e}_1 = (1,0,...)'$, This implies weak convergence of the centered statistic:

$$a_T\Big(\hat{\beta}(\tau)-\beta(\tau)\Big) \xrightarrow{d} -k^{\frac{1}{\alpha}}\mathbf{e}_1 + \operatorname*{argmin}_z Q_\infty(z), \text{ as } \tau T \to k, T \to \infty.$$

We now extend the analysis to tail types 1 and 2, and note that the procedure for these cases is similar, using the following normalized statistics:[4]

$$\begin{array}{lcl} \hat{Z}_T \equiv & a_T\left(\hat{\beta}(\tau)-\beta(q)-b_T\mathbf{e}_1\right), & \text{for type 1 tails,} \\ \hat{Z}_T \equiv & a_T\left(\hat{\beta}(\tau)-\beta(\tfrac{1}{2})\right), & \text{for type 2 tails.} \end{array}$$

The error term $U_t$ and $q$ are defined as in Assumption 1, and the objective functions are written then in the same way as before – (2.1.4), except the integration is taken with respect to the point process:

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

with the canonical constants $(a_T, b_T)$ defined in (1.3.4). The limit point process $\mathbf{N}(\cdot) = \sum_{t=1}^{\infty} 1(\{J_t, X_t\} \in \cdot)$ has points $\{J_i, X_i\}$ distributed as:

$$J_i = \begin{cases} \ln \Gamma_i, & \text{for type 1,} \\ -\Gamma_i^{-\frac{1}{\alpha}}, & \text{for type 2,} \\ \Gamma_i^{\frac{1}{\alpha}}, & \text{for type 3, as already stated,} \end{cases} \tag{2.1.7}$$

and $\{X_i\}$ are i.i.d. with law $F_X$, distributed independently of $\{J_i\}$.

Before we state the result, it would be useful to have a concept of *joint convergence* in distribution of several extremal regression quantiles. We formalize this as follows. Suppose we have $l$ sequences $\{\tau_i, i \le l\}$ s.t $\tau_i T \to k_i$. Index the corresponding to each sequence normalized regression quantile statistics $\hat{Z}_T$ by $k_i$, that is $\hat{Z}_T(k_i)$ (for both $T < \infty$ and $T = \infty$). We show in appendix that $(\hat{Z}_T(k_i), i \le l) \xrightarrow{d} (Z_\infty(k_i), i \le l)$, using the arguments similar to those above.

The following proposition summarizes the findings *under the simplest set of conditions*:

[4] We do not work directly with the centered processes but rather use these forms in order to relate to the point processes in a straightforward way.

***Theorem 1 (Extreme Ranks)***
*Under Assumption 1, as $\tau T \to k, T \to \infty$, $(k = k_1, ..., k_l)$*[5]*:*

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

*where $l(u, v) \equiv 1(u \le v)(v - u)$, and distribution of points $\{J_i, X_i\}$ of $\mathbf{N}$ is defined in (2.1.7). Furthermore, $\big(\hat{Z}_T(k_i), i \le l\big) \xrightarrow{d} \big(Z_\infty(k_i), i \le l\big)$.*

This result implies the weak convergence of the centered statistics

$$\hat{Z}_T^c(k) \equiv a_T \Big( \hat{\beta}(\tau) - \beta(\tau) \Big).$$

***Theorem 2 (Centered Statistics with Extreme Ranks)***
*Under Assumption 1, as $\tau T \to k, T \to \infty$ $(k = k_1, ..., k_l)$*[6]

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

$$\mathbf{c}(k) = \begin{cases} \ln k \ \mathbf{e}_1, & \textit{for type 1,} \\ -k^{\frac{-1}{\alpha}} \mathbf{e}_1, & \textit{for type 2,} \\ k^{\frac{1}{\alpha}} \mathbf{e}_1, & \textit{for type 3,} \end{cases}$$

*Furthermore, $\big(\hat{Z}_T^c(k_i), i \le l\big) \xrightarrow{d} \big(Z_\infty^c(k_i), i \le l\big)$.*

The asymptotic distribution is thus defined by the random variable $Z_\infty$, the optimum of $Q_\infty(\cdot)$. Note that density of $Z_\infty$ can be simulated given the definition of the limit point process $\mathbf{N}$.[7] It also defines the limit density function, which is stated in the next section. Remark 3 shows how this result relates to the classical theory.

**Remark 2.1.1 ( Previous Literature: Extreme Regression Quantiles $(k = 0)$)** The sample extreme or minimal regression quantiles are sample regression quantiles that envelope the data from

[5] For no-covariate case, $k$ should be non-integer to insure a.s. uniqueness of the solution to the stochastic problem.

[6] For no-covariate case, $k$ should be non-integer to insure a.s. uniqueness of the solution to the stochastic problem.

[7] Briefly, the simulation algorithm is as follows: (a) simulate $(J_i, \mathcal{X}_i), i = 1, ..., n$, for $n$ large; (b) solve $\tilde{Z}_{\infty,n} = \text{argmin} \sum_{i=1}^n \rho_{\frac{k}{n}}(J_i - \mathcal{X}_i' z)$; repeat (a) $b$ times. As $b, n \to \infty$, this produces a valid approximation of distribution of $Z_\infty$ by the argument used to show Theorem 1.

below. In single sample they collapse to the first order statistics. Our result contain the extreme regression quantiles as a special case, since extreme quantile case necessarily results from taking $k$: $0 < k < 1$ in near-extreme quantile regressions.[8] Extreme regression quantiles were considered in Smith[80], Portnoy and Jureckova [68], Chernozhukov[17], Knight[48], [49]. Also, Feigin and Resnick [33], using an an estimator that is asymptotically equivalent to the minimal regression quantiles, obtain the limit distributio (in the problem of estimating the autocorrelation coefficient of an AR process with positive i.i.d. innovations). All these results essentially state ( under various levels of generality and regularity conditions) that the limit program is $\max_z \mu_X' z$ s.t. $J_i \geq X_i' z$, $\forall i$. This is simpler than the general case of $k > 0$ given in Theorem 1, as it does not involve stochastic integrations.[9]

## 2.2 Generalization to Congenial Tail Heterogeneity

We propose a model of tail heterogeneity that allows one to flexibly account for possible asymptotic dependence of the tail on the covariate. It may be relaxed to the degree that the key point process $\widehat{\mathbf{N}}$ converges to a well-behaved process $\mathbf{N}$, as required in the appendix.[10] However, this model of tail heterogeneity exhibits certain simplicity and enables to fully characterize limit process $\mathbf{N}$, employing the regular variation theory. This model is also *congenial* from the *prospective of practical inference* – for example, the tail index $\xi$ is retained the same across $X$, scaling constants $(a_T, b_T)$, constants $\mathbf{c}(k)$, and the tail heterogeneity parameters to be introduced are fully characterizable and easily estimable.

For $U_t$ defined in assumption 1, assume the tail equivalence condition:

$$F_{U_t}(z|X_t = x) \sim K(x)F_u(z), \text{ as } z \searrow -\infty \text{ or } z \searrow 0, \tag{2.2.8}$$

uniformly in $x \in \mathbf{X}$, $t$, $F_u$ has type 1-3 tails, $K(\cdot)$ is assumed to be positive continuous function on $\mathbf{X}$, bounded above and away from zero, and normalised so that $K(\mu_X) = 1$.[11]

[8] Koenker and Bassett[6] have shown that the first break-point, defined in terms of $\tau$, is always no less than $1/T$. Hence extreme and all the near-extreme regression quantiles for $k < 1$ are necessarily the same.

[9] It should be said also that the results of this special case are stated for the cases when $U_t$ is independent of $X_t$. We note also that the *special feature* of the extreme regression quantiles is that they can be written as the optimizers of the programming problem: $\max_\beta \bar{X}'\beta$ s.t. $Y_t \geq X_t'\beta, \forall t \leq T$. This property was exploited in the literature to derive the limiting distributions.

[10] Specific examples in these directions are also discussed in the appendix.

[11] If $\mu_X \notin \mathbf{X}$ (e.g. when a covariate is discrete), any other reference point $x^r \in \mathbf{X}$ can be used for normalization: $K(x^r) = 1$.

We deduced that the linearity assumption further implies that:

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

(In case of tail homogeneity, $\mathbf{c} = \mathbf{0}$ in the first case, and $\mathbf{c} = (1, 0, ...)' = \mathbf{e}_1'$ in the others.) We call $\mathbf{c}$ the *tail heterogeneity parameter* or *index*. It measures how strongly the covariates $X_t$ influence tail behavior of error terms $U_t$. Note that $x'\mathbf{c} > 0$ uniformly on $\mathbf{X}$ for types 2 and 3 by assumption.

It is shown that, under conditions (2.2.8) and *weak dependence* conditions on the data sequence, $\widehat{\mathbf{N}}$, constructed as before, with the *canonical scaling constants* defined in terms of $F_u$, weakly converges to $\mathbf{N}$ with points $\{J_i, \mathcal{X}_i\}$ distributed as:

$$(J_i, \mathcal{X}_i) = \begin{cases} (\ln(\Gamma_i) + \mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 1,} \\ (\Gamma_i^{-1/\alpha}\mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 2,} \\ (\Gamma_i^{1/\alpha}\mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 3,} \end{cases} \tag{2.2.9}$$

where $\{\mathcal{X}_i\}$ are i.i.d. with law $F_X$, independently distributed of $\{\Gamma_i\}$, and the centering constants in Theorem 2 are given by

$$\mathbf{c}(k) = \begin{cases} \ln k\mathbf{e}_1 + \mathbf{c}, & \text{for type 1,} \\ -k^{\frac{-1}{\alpha}}\mathbf{c}, & \text{for type 2,} \\ k^{\frac{1}{\alpha}}\mathbf{c}, & \text{for type 3.} \end{cases} \tag{2.2.10}$$

Note that in this case, unlike in the tail-homogeneous case, $J_i, X_i$ are no longer mutually independent. The degree of dependence is determined by $\mathbf{c}$, the tail heterogeneity parameter.

***Theorem 3 (Case with Tail Heterogeneity (and Serial Dependence))***
*With condition (2.2.8) and weak dependence conditions of appendix, replacing (A) and (B) in assumption 1, the statements of Theorems 1 and 2 remain valid, with points $\{J_i, \mathcal{X}_i\}$ of $\mathbf{N}$ defined in (2.2.9) and the centering constants $\mathbf{c}(k)$ defined in (2.2.10).*

**Remark 2.2.1 (Asymptotic Density)** Define $\mathcal{H}$ to be the set of all $d$-element permutations of the integers $1, 2, ....$ Let $\mathcal{X}(h)$ denote the matrix consisting of rows $\mathcal{X}_t, t \in h$, $h \in \mathcal{H}$, and let $J(h)$ denote vector with elements $J_t, t \in h$. It is straightforward to conclude, applying gradient methods of [50] to the limit program, that (i) an argmin can be always expressed as $z = \mathcal{X}(h)^{-1}J(h)$ (as stated, it is a.s. unique if $k$ is not a break-point a.s.) and (ii) $z$ is an optimum iff

$$\zeta_h(z) \equiv (k\mu_X - \sum_{t=1}^{\infty} 1(J_t < \mathcal{X}_t' z)\mathcal{X}_t)\mathcal{X}(h)^{-1} \in \mathcal{D} = [0,1]^d. \tag{2.2.11}$$

$z$ is the unique optimum if $\zeta_h(z) \in \mathcal{D}^o$, the interior of $\mathcal{D}$. Further applying argument of [50], used to state the finite-sample density of regression quantiles, the density of $Z_\infty$ could be written as:

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

where $E$ denotes the expectation over the infinite sequence $\{\mathcal{X}_t\}$.

**Remark 2.2.2 (Relating to Classical Theory)** Note that explicit form of $P$ ( $\zeta_h(z) \in \mathcal{D}^o \mid \{\mathcal{X}_t\}$, $h$ ) is hard to obtain, unless special cases are considered. One such instructive special case is the classical no-covariate theory. We have $X_i = 1$ and $P\left(\zeta_h(z) \in \mathcal{D}^o|h\right) = 1$, if $h = \lceil k \rceil$, 0, otherwise. 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. Note that a.s. uniqueness of $Z_\infty$ as of optimizer of $Q_\infty(\cdot)$ holds if $k$ is non-integer, in view of (2.2.11).

## 2.3 Conclusion

This chapter provided the essential results for the extreme rank regression quantiles in the simplest tail homogeneous regression model. A model of congenial tail heterogeneity was then built to flexibly allow for dependence of tails on covariates. Asymptotic distributions for that model were obtained under i.i.d. and serial dependence conditions.

# Chapter 3

# Asymptotics of Intermediate Ranks

## 3.1 Essential Asymptotics

Intermediate ranks is the case when $\tau \to 0, \quad \tau T \to \infty$. For a given sample, this condition captures the cases when $\tau$ is close to zero, but $\tau T$ (rank) is not small.

In addition to assumption 1, we require existance of density of $F_{u_t}(\cdot|x)$ or, equivalently, of $\partial F_{u_t}^{-1}(\tau|x)/\partial\tau \equiv x'\partial\beta(\tau)/\partial\tau$. Furthermore, we need to strengthen than assumption 1(A) tail-equivalence condition, requiring the derivatives to be tail equivalent: for any $x \in \mathbf{X}$, as $\tau \searrow 0$, uniformly in $t$

$$\partial F_{u_t}^{-1}(\tau|x)/\partial\tau \sim \partial F_u^{-1}(\tau)/\partial\tau \tag{3.1.1}$$

[By linearity this holds uniformly over $\mathbf{X}$]. The added *key* assumption here is that:

$$\partial F_u^{-1}(\tau)/\partial\tau \in \mathcal{R}_{-\xi-1} \tag{3.1.2}$$

where $\mathcal{R}_{-\xi-1}$ is the set of functions regularly varying at 0 with exponent $-\xi - 1$.[1] Dekkers and De Haan[26] have succesfully defended (3.1.2) as a very weak assumption and used it in deriving the limit result under intermediate ranks for the non-regression case, as discussed.[2]

[1] Alternatively, this can stated as: $f_u = \partial F_u(u)/\partial(u)$ is regularly varying of the degree of regular variation of $F$ minus 1; in type 1 cases, $1/f$ is $\Gamma$-varying – see [26].

[2] To see that this assumption holds generally, take for example the cases of the exact algebraic behavior and differentiability near finite or infinite lower end-point: $F_u(z) = z^{-1/\xi}$ as $z \searrow 0$ or $F_u(z) = (-z)^{-1/\xi}$ as $z \searrow -\infty$.

Assumption (3.1.2) has an important implication:[3]

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau f_u(F_u^{-1}(\tau))^{-1}} \to \frac{m^{-\xi}-1}{-\xi}, \forall m > 0, m \neq 1, \text{ as } \tau \searrow 0, \tag{3.1.3}$$

Note that if $\xi = 0$, the quantity on the right hand-side should be interpreted as $\log m$. This property is essential.

In addition to the main result to be stated, it may also be useful to state joint convergence in distribution of several extremal regression quantiles. Fix a reference index (sequence) $\tau$, and then consider $\beta(\tau l_i), i \leq n$. Let $a_T(l) \equiv \sqrt{\tau l T}/|\mu_X' \left(\beta(ml\tau) - \beta(l\tau)\right)|$, for positive $l$ and $m$ ($m \neq 1$). Also let $a_T \equiv a_T$. Define $\mathbf{Z}_T \equiv \Big(a_T(l_i)\big[\hat\beta(l_i\tau) - \beta(l_i\tau)\big], i \leq n\Big)$.

***Theorem 4 (Intermediate Ranks)***
*Suppose Assumption 1, (3.1.1), and (3.1.2) hold, then as $\tau T \to \infty, \tau \searrow 0$*

$$a_T\Big(\hat\beta(\tau) - \beta(\tau)\Big) \xrightarrow{d} N\left(\mathbf{0}, \mathcal{Z}\right), \quad \mathcal{Z} \equiv \mathcal{Q}_X^{-1}\frac{\xi^2}{(m^{-\xi}-1)^2}, \tag{3.1.4}$$

*where $\mathcal{Q}_X \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T EX_tX_t'$. Furthermore, $\mathbf{Z}_T \xrightarrow{d} N(\mathbf{0}, \Omega)$, $\Omega_{ij} \equiv \mathcal{Z}\min(l_i, l_j)/\sqrt{l_i l_j}$. Furthermore, $a_T(l)$ can be replaced by $\sqrt{\tau l T}/\bar X'\big(\hat\beta(ml\tau) - \hat\beta(l\tau)\big)$ without affecting the result.*

**Remark 3.1.1** 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 $\Big(a_T\big(\hat\beta(l_i\tau) - \beta(l_i\tau)\big), i \leq n\Big) \xrightarrow{d} N(\mathbf{0}, \Sigma)$, $\Sigma_{ij} \equiv \Omega_{ij}(l_il_j)^{-\xi}/\sqrt{l_il_j}$.

An outline of the argument now follows. It is brief and differs substantially from the non-regression case. Under simplest conditions, it is meant to *highlight* the role of assumption (3.1.2). Appendix covers general cases.

Since the objective function is not sufficiently smooth, an approach as in the previous section is pursued. The rescaled statistic $\hat Z_T \equiv a_T\big(\hat\beta(\tau) - \beta(\tau)\big)$ solves the following problem: $\min_{z\in\mathbb{R}^d} Q_T(z)$, where

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

[3] See appendix

The goal is to find the finite-dimensional distributional limit $Q_\infty(\cdot)$. Then, using the convexity arguments, conclude that argmin $_z Q_T(z) \xrightarrow{d}$ argmin $_z Q_\infty(z)$. Write

$$\begin{aligned} Q_T(z) \equiv & \left[\frac{-1}{\sqrt{\tau T}}\sum_{t=1}^{T}(\tau - 1[Y_t \leq X_t\beta(\tau)])X_t'z\right] \\ & + \left[\frac{1}{\sqrt{\tau T}}\sum_{t=1}^{T} -\eta_t(z)X_t'z + \eta_t(z)[Y_t - X_t\beta(\tau)]a_T\right] \equiv \; W_T'z \; + \; G_T(z). \end{aligned}$$

where $\eta_t(z) \equiv \big(1(Y_t \leq X_t'\beta(\tau l)) - 1(Y_t \leq X_t'\beta(\tau l) + \frac{X_t'z}{a_T(l)})\big)$.

Appendix shows that definition of $a_T$, (3.1.1), and (3.1.2) *delicately imply*: $E\eta_t(z) = O(f_u(F_u^{-1}(\tau)a_T^{-1})$. Consequently compactness assumption $1(B)$, (3.1.1), and (3.1.3) imply, as $\tau T \to \infty, \tau \searrow 0$:

$$E|G_T(z)|^2 = O(\frac{T}{\tau T}f_u(F_u^{-1}(\tau))a_T^{-1}) = o(1) \tag{3.1.5}$$

Hence $G_T(z) - EG_T(z) \xrightarrow{p} \mathbf{0}$. Appendix shows, applying (3.1.1), (3.1.2), and (3.1.3), that pointwise in $z$

$$E(G_T(z)) \to \frac{1}{2}z'\mathcal{Q}_X z|\frac{m^{-\xi}-1}{-\xi}| \equiv z'\frac{1}{2}J(m)z \tag{3.1.6}$$

Verification of (3.1.6) and (3.1.5) is most difficult task. The rest is simple. By Lindeberg-Feller CLT,

$$W_T \xrightarrow{d} W_\infty \equiv N(\mathbf{0}, \mathcal{Q}_X). \tag{3.1.7}$$

Therefore, the finite-dimensional distributional limit of $Q_T(\cdot)$ is $Q_\infty(\cdot)$:

$$Q_\infty(z) \equiv W_\infty' z + z'\frac{1}{2}J(m)z, \tag{3.1.8}$$

Since $Q_T(\cdot), Q_\infty(\cdot)$ are convex and continuous, and $Q_\infty(\cdot)$ has unique minimizer $Z_\infty \equiv -J^{-1}(m)W_\infty = O_p(1)$, conclude by a convexity theorem (e.g. theorem 5 in [48]) that $Z_T \xrightarrow{d} Z_\infty$. Hence, the first assertion of the theorem is true.

We leave the joint convergence to appendix, and note that the last assertion of the theorem is also true simply by Slutsky and (3.1.4). Indeed, it suffices to show only that:

$$\begin{aligned} \frac{\bar{X}'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \equiv & \frac{\bar{X}'(\hat{\beta}(m\tau) - \beta(m\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} - \frac{\bar{X}'(\hat{\beta}(\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \\ & + \frac{\bar{X}'(\beta(m\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \xrightarrow{p} 1, \end{aligned} \tag{3.1.9}$$

This is so since the first two elements on the r.h.s. are of order $O_P(\frac{1}{\sqrt{\tau T}}) = o_P(1)$.

More generally, we can show

$$\frac{x'(\hat\beta(m\tau) - \hat\beta(\tau))}{x'(\beta(m\tau) - \beta(\tau))} \xrightarrow{p} 1 \text{ uniformly in } x \in \mathbf{X}. \tag{3.1.10}$$

That is, the *empirical regression quantile spacings* $x'(\hat\beta(m\tau) - \hat\beta(\tau))$ *behave approximately like* the *population quantile spacings* $x'(\beta(m\tau) - \beta(\tau))$, as $\tau T \to \infty, \tau \searrow 0$. This *remarkable* property allows a very simple construction of various estimators of tail index $\xi$ and tail heterogeneity parameter $\mathbf{c}$ in the model of congenial tail heterogeneity.

**Remark 3.1.2** **(On Rates of Convergence and Empirical Regression Quantile Spacings.)** Note that (3.1.9) *does not simply follow from consistency* of the estimator under intermediate ranks, since by (3.1.4) the latter is not true unless:

$$\sqrt{\tau T}/\mu_X'(\beta(m\tau) - \beta(\tau)) \to \infty. \tag{3.1.11}$$

(3.1.11) may sometime be violated for type 2 and type 1 cases (still (3.1.9) remains true). Indeed, take the case of $F_u$ 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 sequence of $\tau$ is s.t. $\tau = T^{-\lambda}$, $\lambda \in (0, 1)$, and if $\frac{2\lambda}{1-\lambda} > \alpha > 0$, then $\sqrt{\tau T}/\mu_X'(\beta(m\tau) - \beta(\tau)) \to 0$, implying divergence. That is, divergence may arise when when (i) $\tau$ approaches 0 at a sufficiently fast rate and (ii) the tails are sufficiently thick. Notwithstanding, (3.1.10) remains valid.

The reader may observe that the weak convergence results obtained remain valid under dependence and heterogeneity as long as (i) a suitable CLT, (ii) LLN, and (iii) analytical properties of the kind (3.1.2) apply. This generalization is formalized in the appendix. Next section gives a sufficient rather general case – the model of congenial tail heterogeneity and a weak dependence condition.

## 3.2 Generalizations to Congenial Tail Heterogeneity

The model of congenial tail heterogeneity is introduced and discussed in section 3 (and appendix). Assumptions, replacing (3.1.1) and (3.1.2), are stated as follows. For all $x \in \mathbf{X}$, as $\tau \searrow 0$:

$$\begin{gathered} \partial F_{u_t}^{-1}(\tau|x)/\partial\tau \equiv x'\partial\beta(\tau)/\partial\tau \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau, \text{ and} \\ \partial F_u^{-1}(\tau)/\partial\tau \in \mathcal{R}_{-\xi-1} \end{gathered} \tag{3.2.12}$$

In addition to tail heterogeneity, we may also allow for serial dependence. Appendix gives a strong mixing and a no-clustering condition that makes the asymptotic distribution not affected by dependence. This is expected in view of localization. In the unrelated literature on kernel regression (e.g. [73], [8]), localization of CLTs and LLNs has the same result on the limit distribution.

***Theorem 5 (Case with Tail Heterogeneity and Serial Dependence)***
*Suppose (3.2.12) holds, replacing assumption 1 (where appropriate). Assume $\{Y_t, X_t\}$ is independent or (dependent with conditions in appendix), all results of the theorem 4 remain valid, except variance*

$$\mathcal{Z} \equiv \mathcal{Q}_{\mathcal{H}}^{-1} \mathcal{Q}_X \mathcal{Q}_{\mathcal{H}}^{-1} \frac{\xi^2}{(m^{-\xi} - 1)^2}$$

$\mathcal{Q}_{\mathcal{H}} \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E1/\mathcal{H}(X_t) X_t X_t'$, *and*

$$\mathcal{H}(x) = \begin{cases} 1 & \text{for type 1,} \\ (x'\mathbf{c}) & \text{for type 2 and 3 tails.} \end{cases}$$

*Joint convergence remains valid, with $\Omega_{ij}$ equal to $\mathcal{Z} min(l_i, l_j)/\sqrt{l_i l_j}$.*

Remarks 4 and 5 apply in this case too. Note that for type 2 and 3 tails $x'\mathbf{c}$ is bounded above and away from zero uniformly on $\mathbf{X}$, by construction.

## 3.3 Conclusion

This chapter provided the essential results on the intermediate rank regression quantiles in the simplest tail homogeneous regression model. A model of congenial tail heterogeneity was then built to flexibly allow for dependence of tails on covariates. Asymptotic distributions for that model were obtained under i.i.d. and serial dependence conditions.

# Chapter 4

# Extremal Processes and Related Inference

## 4.1 Introduction

This chapter develops a theory of extremal quantile regression process and related inference processes.

> We formulate Extremal QR process by taking all sample quantile regression statistics $\hat{\beta}(\tau)$ with values of quantile index $\tau$ in a near-extreme range (close to zero).
>
> Inference processes and resulting Kolmogorov-Smirnov and Cramer-Von Mises test statistics are defined as function(al)s of such processes.
>
> Asymptotic behavior of the extremal quantile and inference processes, as well as of the resulting test statistics, is obtained under conditions reflecting rare nature of the conditional extremal events [ the extreme and intermediate rank].

Such *formulations* and *asymptotic theory* are *novel.* See [52] for the theory of central, non-extremal quantile processes and the inference on such.

The asymptotic theory of extremal quantile regression processes and the resulting from it inference theory *encompasses both the pointwise and global inference about the extremal quantile regression model.* As a result, the applications are plentiful. For example, in the analysis of high market risk (value-at-risk), it is important to check if a conditional model differs from the unconditional one; or to check which covariates matter most for high level risk. See [20] for an empirical work in this direction. Other examples include problems of insurance, search economics, dynamic option pricing, and other extreme event analyses where conditioning is crucial.

A further set of statistical applications includes estimation of tail index, tail heterogeneity index, and of high conditional quantiles by extrapolation, as developed in the next chapter. There we define the requisite estimators from the extremal quantile regression process (that is, from many values of the extremal QR statistics). The asymptotics is a consequence of this chapter.

The organization of this chapter is as follows. In section 4.2 extremal quantile regression processes are defined and embedded in rich functional metric spaces. The asymptotic behavior under data scarcity (extreme and intermediate rank conditions) is established in sections 4.3 and 4.4. Inference processes and resulting test statistics are defined and studied in section 4.5.

## 4.2 Definitions

The *setting* assumed throughout is the *model of the congenial tail heterogeneity.* The basic benchmark model with i.i.d. errors and homogeneous tails is subsumed as a special case. Notwithstanding, the results are new for the most basic model.

Near-extreme sample regression quantiles can be indexed by a location $l$ relative to a reference quantile $\tau$, $\hat{\beta}(\tau l)$. A process then is formed by varying $l$ in some index set $\mathcal{L}$. Formally we define the following processes:

For **extreme ranks**:

$$Z_T^c(\cdot) \equiv \{Z_T(k), k \in \mathcal{K}\}$$

where $\mathcal{K} = [\underline{k}, \bar{k}]$, and $Z_T^c(k) = a_T(\hat{\beta}(k/T) - \beta(k/T))$ is the properly scaled statistic as defined in chapter 2 (superscript "c" stands for centering). Correspondingly, $Z_T(\cdot)$ denotes the un-centered process with coordinates $Z_T(k)$. At the break points, when $Z_T^c(k)$ is not uniquely defined as an optimizer of the RQ problem, the ties are resolved so that the resulting process $Z_T^c(\cdot)$ is *cadlag.* Number of break-points in set $\mathcal{K}$ is finite for $T < \infty$ and is stochastically bounded for $T = \infty$. A conjectured *limit process* is the finite-dimensional limit

$$Z_\infty^c(\cdot) \equiv \{Z_\infty^c(k), k \in \mathcal{K}\}$$

where $Z_\infty = Z_\infty(k) - c(k)$, as defined in chapter 2.

For **intermediate ranks**:

$$Z_T(\cdot) \equiv \{Z_T(l), l \in \mathcal{L}\}$$

where $\mathcal{L} = [\underline{l}, \bar{l}]$, and $Z_T(l) = a_T(\hat{\beta}(l\tau) - \beta(l\tau))$ is the properly scaled statistic as defined in chapter 3. In case when $Z_T(l)$ is not uniquely defined as an optimizer of the RQ problem, the ties are resolved so that the resulting process $Z_T(\cdot)$ is *cadlag*. For $T < \infty$ number of break-points is finite. The limit process will have no break-points a.s., for the conjectured limit is a Gaussian process with continuous sample paths:

$$Z_\infty(\cdot) \equiv \{Z_\infty(l), l \in \mathcal{L}\},$$

where $(Z_\infty(l_i), i \leq n)$ is a Gaussian vector whose distribution is defined in chapter 3.

With the conjectured limits above, there are few important remarks in order. We like to pick a metric space for our processes so that the resulting set of continuous functionals is as rich as possible. Indeed, inference processes and test statistics will be constructed as continuous functionals. The larger the set is, the richer statistics we can get. One most attractive functional space is the space of bounded functions on the index set $\mathcal{K}$ or $\mathcal{L}$ equipped with the (strongest) uniform metric [$\ell^\infty(\mathcal{K})$ and $\ell^\infty(\mathcal{L})$]. Under extreme ranks the limit is a stochastic cadlag function, so there is a measurability problem ($Z_T(\cdot)$ is not Borel-measurable as a map from the sample space to $\ell^\infty(\mathcal{L})$). So we must give up, and pick another space. Our choice is $L^2(\mathcal{K}, \mu_L)$ – the space of functions mapping $\mathcal{K}$ to $\mathbb{R}^d$ integrable w.r.t. Lebesgue measure. The limit process is shown to be measurable, and the resulting collection of continuous functional is sufficiently rich. Under intermediate ranks the limit $Z_\infty(\cdot)$ will be shown (luckily) measurable [ although $Z_T(\cdot)$ is not (for it is a cadlag)]. Thus using the weak convergence theory in ([85]) will be fruitful, and the set of continuous functionals will be maximal.

## 4.3 Behavior of Extremal RQ Processes with Extreme Ranks

The basic settings are those of chapter 2.

***Assumption 2***
*The following hold:*

- *Assumptions of Theorems 1 or, more generally, Theorem 3 of chapter 2.*

- *In addition, for type 2 tails*

$$\sup_{x \in \mathbf{X}, k \in \mathcal{K}} x' Z_T(k) < 0 \ w.p. \ \to 1, \ for \ T \leq \infty. \tag{4.3.1}$$

**Remark 4.3.1** The added condition(4.3.1) holds trivially when $d = 1$, and otherwise when covariates are discrete (which for all practical purposes may suffice). Extending it to general situations seems to be a technical matter. It helps us to guarantee that number of break-points for type 2 tails is stochastically bounded.

Here is the essence of the proof. Working directly with $Z_T(\cdot)$ appears to be difficult, since $Z_T(\cdot)$ is a piece-wise constant function with randomly spaced break-points. Number of break-points, $\#\mathcal{K}_{bT}$ is stochastic albeit stochastically bounded. This property allows approximating $Z_T(\cdot)$ asymptotically arbitrarily well by a well-behaved piece-wise constant approximation map $AZ_T(\cdot)$ with a finite number of regularly spaced break-points. Map $AZ_T(\cdot)$ is defined by a finite number of projections $\{Z_T(k_i), i \leq n\}$. The weak convergence is now entirely a consequence of finite dimensional convergence [ Theorem 1 for the benchmark model, and Theorem 3 for the congenial tail heterogeneity model in Chapter 2.]

***Lemma 4.3.1***
*Centering constants $\mathbf{c}(k)$ may be obtained as limits uniformly in $k \in \mathcal{K}$.*

Proof is in the appendix for chapter 2 (see equation (A.4.35)).

***Theorem 6 (Extremal RQ Process with Extreme Ranks)***
*Given assumption 7, $Z_T(\cdot) \Rightarrow Z_\infty(\cdot)$ in $L_2(\mathcal{K}, \mu_L)$, where $Z_\infty(k) = \text{ argmin }_z \ [-k\mu_X' z + \int l(u, x'z) d\mathbf{N}(u, x)]$, $l(u, v) \equiv 1(u \leq v)(v - u)$, and $\mathbf{N}$ is s Poisson point process with mean measure that depends on tail type as defined in Theorem 1 and 3. $Z_T^c(\cdot) \Rightarrow Z_\infty(\cdot) - \mathbf{c}(\cdot)$ in $L_2(\mathcal{K}, \mu_L)$, where $\mathbf{c}(k)$ is defined in Theorems 2 and 4 of Chapter 2.*

This theorem is a corollary of the appendix.

Appendix generalizes this result by requiring only certain key conditions to hold.

## 4.4 Behavior of Extremal RQ Process with Intermediate Rank

We like to study the convergence of map: $Z_T(\cdot) = \{a_T(l)(\hat{\beta}(l\tau) - \beta(l\tau)), l \in \mathcal{L}\}$. The choice of metric space $\ell^\infty(\mathcal{L})$ is attractive. The uniform metric on $\ell^\infty(\mathcal{L})$ allows a maximal set of continuous functionals. This is a convenient property, since the main use of the processes is sequential: inference processes and tail index and other estimators are defined as continuous functionals of process $Z_T(\cdot)$.

***Theorem 7 (Extreme Process with Intermediate Ranks)***
*Under conditions of Theorem 4 and 5 in Chapter 3, as $\tau \to 0, \tau T \to \infty$*

$$Z_T(\cdot) \Rightarrow Z_\infty(\cdot) \quad \text{in } \ell^\infty(\mathcal{L}),$$

*where $Z_\infty(\cdot)$ is a centered Gaussian process with covariance $E(Z_\infty(l_i)Z_\infty(l_j)') = \Omega_{ij}$ stated in Theorems 4 and 5 of Chapter 3. $Z_\infty(\cdot)$ has a.s. continuous paths.*

## 4.5 Inference Process

### 4.5.1 Inference Processes under Extreme Ranks

We would like to check the hypotheses about the true near-extreme regression quantile coefficients $\beta(\tau)$. In this section we do so formally under the extreme ranks condition: $\tau \searrow 0, \tau T \to k$. This form of asymptotic analysis is simply an approximation of practical situations with given large sample size $T$ yet small rank $\tau T$. Hence consider the basic hypothesis:

$$H_0 : R(k)'\beta(\tau) - c(\tau) = 0, [\text{ as } \tau T \to k, T \to \infty],$$

where $R(k)$ is some bounded matrix that depends on rank $k$ and *has a full column rank.*

To test the $H_0$ for a given $k$ [ or quantile $\tau = k/T$ ], consider a "Wald" statistic

$$W_T(k) = a_T\big(R(k)'\hat{\beta}(\tau) - c(\tau)\big)'\widetilde{\Omega}_1(k)a_T\big(R(k)'\hat{\beta}(\tau) - c(\tau)\big),$$

where $\widetilde{\Omega}_1(k)$ is some conformable matrix converging in probability to $\Omega_1(k)$, a fixed positive definite matrix. Then under $H_0$:

$$W_T(k) \overset{H_0}{=} a_T\big(\hat{\beta}(\tau) - \beta(\tau)\big)'R(k)\widetilde{\Omega}_1(k)R(k)'\big(\hat{\beta}(\tau) - \beta(\tau)\big)a_T.$$

Hence under $H_0$

$$W_T(k) \xrightarrow{d} Z^c_\infty(k)'R(k)\Omega_1(k)R(k)'Z^c_\infty(k). \tag{4.5.2}$$

It is useful now to consider behavior under contiguous ("local") alternatives:

$$H_{aT} : R(k)'\beta(\tau) - c(\tau) = \frac{\delta(k)}{a_T}, [\text{ as } \tau T \to k, T \to \infty].$$

Under sequence of alternatives $\{H_{aT}\}$:

$$W_T(k) \xrightarrow{d} \big(Z_\infty^c(k)'R(k) + \delta(k)'\big)\Omega_1(k)\big(R(k)'Z_\infty^c(k) + \delta(k)\big). \tag{4.5.3}$$

Note that

- if $a_T \to \infty$ [type 3 tails and some type 1 tails], test is asymptotically consistent, that is, $H_0$ is rejected under a fixed alternative w.p. $\to 1$.
- if $a_T \to 0$ [type 2 and some type 1 tails] the test is inconsistent, that is $H_0$ is *not* rejected under a fixed alternative w.p. $\to 1$. This has a finite sample flavor. *Is test then sensible at all?*
- *Yes it is.* Behavior under local alternatives (4.5.3) describes the test power *meaningfully.* Fix $T$ large, then the test statistic is approximately distributed as the r.h.s. of (4.5.3). The statistic is a quadratic form of $|\delta(k)|$ measuring departure from the null. The larger is $|\delta(k)|$, the larger is the statistic. Hence for sufficiently large $|\delta(k)|$, the probability of rejecting $H_0$ is arbitrarily close to 1.

The test is thus *quite sensible* for any practical setting with fixed large sample size and small rank $\tau T$. The right- hand-side of (4.5.2) describes its distribution by a random variable $Z_\infty^c(\cdot)$ defined from a point process $\mathbf{N}$ [that deepens on tail and underlying nuisance parameters]. Its distribution, albeit not tabulated, can be easily simulated – Chapter 6.

Finally, the results developed earlier allow to consider the simultaneous tests about $\beta(\tau)$ with $\tau$ in near-extreme ranges. Such ranges are indexed by rank $k$. That is, formally we look at ranges $\tau$ s.t. $\tau T \to k, k \in \mathcal{K}$.

To address this testing problem, we formulate the inference process $W_T(\cdot)$ by taking values of the Wald statistic for ranks $k$ in the set $\mathcal{K}$:

$$W_T(\cdot) = \{W_T(k), k \in \mathcal{K}\}$$

Let us now embed the *Wald inference process* into $L_2(\mathcal{K}, \mu_L)$. Repeated use of the continuous mapping theorem yields the requisite result.

***Corollary 7.1 (Inference Processes under Extreme Ranks)***
*Under conditions of Theorem 6, if* $\widetilde{\Omega}_1(\cdot) \xrightarrow{p} \Omega_1(\cdot)$ *uniformly in* $\mathcal{K}$, *and* $\Omega_1(\cdot)$, $\delta(\cdot)$, $R(\cdot)$ *are continuous and bounded*

$$W_T(\cdot) \Rightarrow W_\infty\big(Z_\infty^c(\cdot)'R(\cdot) + \delta(\cdot)'\big)\Omega_1(\cdot)\big(R(\cdot)'Z_\infty^c(\cdot) + \delta(\cdot)\big),$$

*in $L_2(\mathcal{K}, \mu_L)$ under sequence of alternatives $\{H_{aT}\}$. Consequently, the Cramer-Von Mises statistic, defined for any $h(\cdot)$ in $L_2(\mathcal{K}, \mu_L)$ as $\int_{\mathcal{K}} h(k)W_T(k)dk \xrightarrow{d} \int_{\mathcal{K}} h(k)W_\infty(k)dk$.*

A simple version of the Cramer-Von Mises results by setting $h(k) = 1$. Then if for a (Lebesgue measure) positive range of values $k$ [or corresponding ranges of $\tau = k/T$] there is a significant departure from null $\delta(k)$, the test will have a non-trivial power (under $H_{aT}$). Thus the earlier power analysis applies here in much the same sense.

Can we go beyond the Cramer-Von Mises type statistic? No, unless we want to change the metric space. Indeed, Cramer-Von Misses statistics coincide with the set of all continuous functionals from $L_2(\mathcal{K}, \mu_L)$ to the real line.

Kolmogorov-Smirnov test statistics of the form $\sup_k |W_T(k)|$ are *ruled out* under extreme ranks, because they are not continuous functionals from the metric space $L_2(\mathcal{K}, \mu_L)$. Additionally, measurability is at stake in this case.

The test statistic above can be easily computed exactly, since its a function of the extremal quantile process which is by itself a cadlag with a finite number of break-points. The limiting distribution is either to be simulated or approximated by subsampling.

### 4.5.2 Inference Processes under Intermediate ranks

In empirical applications it is imperative to check the hypotheses about true near-extreme regression quantile coefficients $\beta(\tau)$. In this section we develop a formal theory to carry out such tests under the intermediate rank condition: $\tau \searrow 0, \tau T \to \infty$. This form of asymptotic analysis is an approximation of practical situations with given large sample size $T$, near-extreme quantile of interest $\tau$, yet not very small rank $\tau T$. Thus formally we consider the basic hypothesis:

$$H_0 : R(l)'\beta(l\tau) - c(\tau l) = 0, [\text{ as } \tau T \to \infty, \tau \searrow 0, ]\forall l \in \mathcal{L},$$

where $R(l)$ is some bounded continuous in $l$ matrix. Fix $l$ for a moment.

To test the above hypothesis for a given $l$, first consider a simple version of a Wald statistic:

$$W_T(l) = a_T(l)\big(R(l)'\hat{\beta}(l\tau) - c(l\tau)\big)'\widetilde{\Omega}_2(l)a_T(l)\big(R(l)'\hat{\beta}(l\tau) - c(l\tau)\big),$$

where $\widetilde{\Omega}_2(l)$ is some conformable matrix uniformly converging in probability to $\Omega_2(l)$, a fixed continuous and uniformly p.d. in $l$ matrix. Then under $H_0$

$$W_T(l) \stackrel{H_0}{=} a_T(l)\big(\hat{\beta}(l\tau) - \beta(l\tau)\big)'R(l)\widetilde{\Omega}_2(l)a_T(l)R(l)'\big(\hat{\beta}(l\tau) - \beta(l\tau)\big).$$

Hence

$$W_T(l) \xrightarrow{d} R(l)'Z_\infty(l)\Omega_2(l)R(l)'Z_\infty(l)$$

where $(Z_\infty(l_i), i \leq n)$ is the random Gaussian vector with distribution defined in Theorem 5 in Chapter 3 (Theorem 4 for the benchmark model).

It is useful next to consider the behavior of the test statistic under contiguous ("local") alternatives:

$$H_{aT} : R(l)'\beta(l\tau) - c(l\tau) = \frac{\delta(l)}{a_T(l)}, \text{ as } \tau T \to \infty, \tau \searrow 0.$$

By Theorem 5 in chapter 3, under $H_{aT}$

$$W_T(l) \xrightarrow{\mathcal{D}} \big(Z_\infty(l)'R(l) + \delta(l)'\big)\Omega_2(l)\big(R(l)'Z_\infty(l) + \delta(l)\big). \tag{4.5.4}$$

Next consider test about $\beta(\tau)$ for the ranges of values of $\tau$ simultaneously. For that purpose construct the *Wald inference process* as:

$$W_T(\cdot) = \{W_T(l), l \in \mathcal{L}\}$$

Next step is to discover the behavior $W_T(\cdot)$ in metric space $\ell^\infty(\mathcal{L})$. It is easy by virtue of the continuous mapping theorem and Theorem 2 of this chapter.

***Corollary 7.2 (Inference Processes with Intermediate Rank)***

*Given stated assumptions and conditions of Theorem 7*

$$W_T(\cdot) \Rightarrow W_\infty(\cdot) \equiv \big(Z_\infty(\cdot)'R(\cdot) + \delta(\cdot)'\big)\Omega_2(\cdot)\big(R(\cdot)'Z_\infty(\cdot) + \delta(\cdot)\big)$$

*in $\ell^\infty(\mathcal{L})$, under $H_{aT}$. $Z_\infty(\cdot)$ is the Gaussian process defined in Theorem 7. Consequently, for any integrable $h$, Cramer-Von Misses statistic $\int_{\mathcal{L}} W_T(l)h(l)dl$ converges weakly to $\int_{\mathcal{L}} W_\infty(l)h(l)dl$; and Kolmogorov-Smirnov statistic $\sup_{l\in\mathcal{L}} W_T(l)$ converges weakly to $\sup_{l\in\mathcal{L}} W_\infty(l)$.*

Local power analysis of the previous section applies here along the same lines. It should be said that the test statistics in the theorem are computable exactly. The processes can also be approximated arbitrarily well by a finite sub-net $\{W_T(l_i), i \leq n\}$, which gives an alternative way to easily compute the test statistic. The limit distributions will typically have to be simulated. Alternatively, bootstrap or subsampling can be used to compute the critical values. Properties of bootstrap approximation are yet to be established however. On the other hand, since the limit inference process is itself approximable by a finite sub-net (by tightness), validity of subsampling bootstrap is now a trivial consequence of its validity for approximating the finite-dimensional limit distribution.

## 4.6 Conclusion

We defined Extremal QR process by taking all sample quantile regression statistics $\hat{\beta}(\tau)$ with values of quantile index $\tau$ in a near-extreme range (close to zero). Inference processes and resulting Kolmogorov-Smirnov and Cramer-Von Mises test statistics are defined as function(al)s of such processes. Asymptotic behavior of the extremal quantile and inference processes, as well as of the resulting test statistics, is obtained under conditions reflecting rare nature of the conditional extremal events. These conditions are extreme and intermediate ranks. A further set of statistical applications includes estimation of tail index, tail heterogeneity index, and of high conditional quantiles by extrapolation. The requisite estimators could be defined from the extremal quantile regression process (that is, from many values of the extremal QR statistics). The asymptotics is a consequence of this chapter.

# Chapter 5

# Inference

## 5.1 Introduction

The aim of this chapter is to provide inference tools to apply our theory to data. Also we demonstrate through Monte-Carlo work that this theory can be valuable in practice. Accordingly, this chapter is organized as follows.

In sections 5.2 and 5.3 we build a novel theory of inference for the *conditional extremal quantiles*. Section 5.2 builds estimators of the tail index and the tail heterogeneity index in the linear quantile regression model. Our estimators are based on the spacings of the sample extremal regression quantiles. These estimators are *completely novel*. Section 5.3 then describes several confidence interval methods: analytical and resampling constructions. Relevant results on consistency are stated.

In section 5.4. we exhibit the essence of Monte Carlo work contained in Appendix C and [19]. This Monte-Carlo strongly supports the theory developed in the previous and current chapters.

## 5.2 Inference about Tails

The inference about tails is an important part of statistical analysis. Methods of the tail index estimation have been developing since 1930s. Significant contributions have been made by [40], [61], [26] and others. Currently this research area remains a very active one (see [31]). Here we build estimators of the tail index and the tail heterogeneity index in the linear quantile regression model with the congenial tails. Our estimators are based on the spacings of the sample extremal regression quantiles. These estimators are consistent and *novel*, and they enjoy a good finite sample performance ([19]). Thus, for the first time in literature, to our knowledge, we offer *estimators of tail index and tail heterogeneity index* (parameter we introduced in Chapter 2) *based on regression quantiles*.

### 5.2.1 The Regression Quantile Tail Index Estimator

Our construction is in the spirit of Pickands[61]. We have discovered in Chapter 3 that the spacings of the intermediate rank sample regression quantiles can consistently approximate the population spacings. Since ratios of the population spacing can reveal the information about tail index and the tail heterogeneity index, so can their consistent empirical analogues.

Indeed, it is the property of the linear quantile regression model with the congenial tails that:

$$\nu_\tau(m,l,s,x) \equiv \frac{x'(\beta(mls\tau)-\beta(ls\tau))}{x'(\beta(ms\tau)-\beta(s\tau))} \to l^{-\xi} \text{ as } \tau \searrow 0 \ , \tag{5.2.1}$$

for some collection of fixed indices $l$, $m$, and $s$.

Denote by $M$ any compact subset of $(0,\infty)$ that excludes 1, and by $\mathcal{L}$ – any compact subset of $(0,\infty)$. Then uniformly for all such indices $(m,l,s,x)$ in $\Theta \equiv M \times M \times \mathcal{L} \times \mathbf{X}$,

$$\nu_\tau(m,l,s,x) \to l^{-\xi} \text{ as } \tau \searrow 0 \ , \tag{5.2.2}$$

by Lemma A.4.1. Next define

$$r_\tau(m,l,s,x) \equiv -\ln \nu_\tau(m,l,s,x)/lnl.$$

Denoting by $\theta$ arguments $(m,l,s,x)$ it will be convenient to abbreviate the l.h.s. as $r_\tau(\theta)$. We always restrict $\theta$ to compact set $\Theta$. To indicate dependence of $\theta$ on an argument, we will write, for example, $\theta_l$. Thus uniformly in $\Theta$

$$r_\tau(\theta_l)/lnl \to \xi.$$

In this construction we call $\tau$ – a reference quantile, and $s$ – a *location index* relating all other indices $\tau'$ to $\tau$. $l$ and $m$ are two important tuning parameters:

- $l$ is a *shift index* relating the quantile spacing in the numerator to that in the denominator $(0 < l \neq 1)$.
- $m$ is a *spacing index* determining the size of the quantile spacing $(0 < m \neq 1)$.

Ratio of the shifted conditional quantile spacings reveals information about the tail index (as $\tau \searrow 0$.) Varying all these parameters $\theta$ gives us a *variety* of ratios of spacings. Their *empirical analogues* constitute a *variety of estimators*, optimal or (perhaps reasonable) combinations of which can be chosen *adaptively*.

Now define the empirical regression analogues:

$$\hat{\nu}_\tau(\theta) \equiv \frac{x'(\hat{\beta}(mls\tau) - \hat{\beta}(ls\tau))}{x'(\hat{\beta}(ms\tau) - \hat{\beta}(s\tau))},$$
$$\hat{r}_\tau(\theta) \equiv -\ln \hat{\nu}_\tau(\theta_l)/lnl.$$

We call $r_\tau(\cdot)$ the population log-quantile spacings function or simply $r$-function, and and $\hat{r}_\tau(\cdot)$ – the empirical log-quantile spacing process or simply $\hat{r}$-process.

***Theorem 8 (Tail Index Estimation)***
*Given assumptions of Theorem 5, for any $\theta$ in $\Theta$, as $\tau T \to \infty, \tau \searrow 0$*

$$\hat{r}_\tau(\theta) \xrightarrow{p} \xi.$$

*Given assumptions of Theorem 7, as $\tau T \to \infty, \tau \searrow 0$*

$$\hat{r}_\tau(\theta) \xrightarrow{p} \xi \text{ uniformly in } \Theta.$$

Since (5.2.2) holds uniformly in $\Theta$ by Lemma A.4.1, proof of this theorem only involves showing that pointwise or uniformly in $\Theta$ the empirical regression quantile spacings can consistently replace the population quantile spacings (Appendix C).

The first result of the theorem states that given any value of $\theta$, the estimator $\hat{r}_\tau(\theta)$ is consistent. The second result strengthens this to uniform consistency. Therefore, all kinds of convex combinations and other transforms of $\{\hat{r}_\tau(\theta), \theta \in \Theta\}$ are consistent for $\xi$. Now we have a variety of estimators to choose from.

*Choice*

Which tuning parameters should one pick? In order to understand better the issues involved, consider a simulation example. This example involves a simple model $Y_t = X_t'\beta_0 + U_t$, where $U_t$ is standard normal, and sample size is 500. The true value of tail index $\xi$ is 0. Depicted in Figure 5.1 is the estimator $\hat{r}_\tau(m, l, \bar{X})$ function for different values of tuning parameters $m$ and $l$ and $\tau$. Note that since, for a given sample size, varying $s$ or $\tau$ is equivalent, we fixed $s = 1$ and suppressed it in notation.

In the figure, we want to identify a set of quantile indices $\tau$ where the estimate is reasonable. Unconditional extreme value theory recommends to use the so-called Hill plots that depict $\tau$ or rank $k = \tau T$ against the estimator $\hat{r}_\tau$. The "right" value of $\tau$ is chosen to correspond to a flat or stable area of the plot. Indeed, if all estimators $\hat{r}_\tau$ are consistent, their values should be close to each other. This implies that there should exist a relatively stable area of the plot $(\tau, \hat{r}_\tau)$.

Once we locate such a region of flatness, it seems very sensible to average over all $\hat{r}_\tau$ in that area. It should be stressed, however, that the optimal weighting in this averaging operation is not obtainable from the consistency results above. We need to know the asymptotic distribution of the $\hat{r}$-process to calculate the optimal weights. The asymptotic distribution theory comes at the cost of additional strong assumptions that impose conditions on the rate at which $\tau T \to \infty$ and some higher order regular variation properties ([18]). Such assumptions are rarely verifiable in practice, but the theory may be used as an approximate guide to choose sensible combinations of $\{\hat{r}(\theta), \theta \in \Theta\}$. A simple strategy we used so far in practice is to simply integrate or take the median over the stable region of the plot ([20]).

If we look at Figure 5.1, however, we see no such stable areas, except on the plot in the top right corner. Figure 5.1 clearly demonstrates that the closer are the tuning parameters $m$ and $l$ to 1, the more volatile is the $\hat{r}$-function, and vice versa. Given the above consistency principle, choosing the tuning parameters $(l, m)$ to produce stable $\hat{r}$ functions seems reasonable. In Figure 5.1 this occurs for $l = .5$ and $m = \{1.5, 2\}$ (top right corner). Integrating or taking the median value over the stable region (from $\tau = .5$ to $\tau = .15$) produces a rather sharp estimate of the tail index $\xi$. Further Monte-Carlo examples in [19] suggest that such principle works well.

To summarize, an appropriate procedure of choosing the estimator consists of two steps:

- choose values $(l^*, m^*)$ that generate stable regions $(\tau, \hat{r}_\tau(l^*, m^*))$,

- aggregate over all estimators $\{\hat{r}_\tau(l^*, m^*)\}$ in that stable region (e.g. compute average, median etc.)

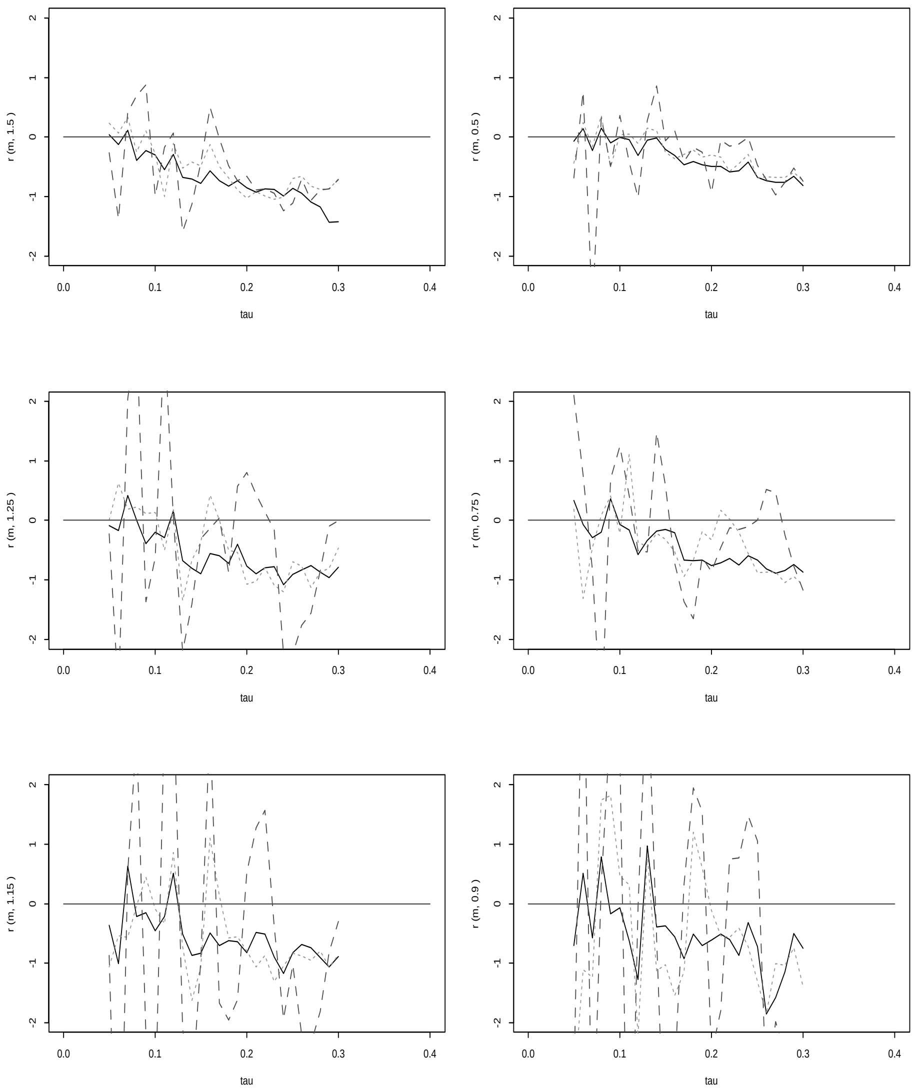


Figure 5.1: An Example of Tail Index estimation. The depicted is the estimator (function) $\hat{r}_\tau(m, l, \bar{X})$ for different values of $m$ and $l$ and $\tau$. On every plot '—' line corresponds to setting constant $m = 2$, '- - -' – $m = 1.5$ and '— — —' – $m = 1.2$. Plots correspond (from left to right, then down) to values of $l$ equal to 1.5, .5, 1.25, .75, 1.15, .9.

### 5.2.2 Estimation of the Tail Heterogeneity Index

An important parameter of the congenial tail heterogeneity model is $\mathbf{c}$. Assume that $\xi \neq 0$ (for all practical purposes). Let us examine the ratio of quantile spacings evaluated at two different values of regressors $x$: $x = x$ and $x = \mu_X$ ($\mu_X = \text{plim}\bar{X}$). As $\tau \searrow 0$

$$\mu(x, m, s) = \frac{x'(\beta(ms\tau) - \beta(s\tau))}{\mu'_X(\beta(ms\tau) - \beta(s\tau))} \longrightarrow x'\mathbf{c}.$$

The empirical analogue $\hat{\mu}(x, m, s)$, constructed from the sample regression quantiles, is therefore a consistent estimator of the linear index $x'\mathbf{c}$.

***Theorem 9 (Tail Heterogeneity Index Estimation)***
*Suppose assumptions of Theorem 5 or Theorem 7 are true and $\xi \neq 0$, then( respectively) pointwise or uniformly in $(m, s, x) \in M \times \mathcal{L} \times \mathbf{X}$, as $\tau T \to \infty, \tau \searrow 0$*

$$\hat{\mu}(x, m, s) = \frac{x'(\hat{\beta}(ms\tau) - \hat{\beta}(s\tau))}{\bar{X}'(\hat{\beta}(ms\tau) - \hat{\beta}(s\tau))} \xrightarrow{p} x'\mathbf{c}$$

Proof of this result involves the same logic as the previous theorem. An obvious consequence of Theorem 9 is that a consistent estimator $\hat{\mathbf{c}}$ can be constructed by minimizing distance (e.g. by least squares) between $\{\hat{\mu}(x_j, m_i, s_k)\}$ and $\{x'_j c\}$ for a sufficiently rich, possibly finite collection of values $x_j, m_j, s_k$. The minimal requirement is that collection $\{x_j, j \geq 1\}$ should span $\mathbb{R}^d$. So only $d$-points may suffice. Discussion of the tuning parameter choice in the previous section applies here equally well.

## 5.3 Confidence Intervals

### 5.3.1 Inference under Extreme Ranks

Ignoring, at first, how we obtain the approximation of the limiting distribution (either through asymptotic means or resampling), we note there are several ways to construct confidence intervals. The problem is to define the best intervals. Unlike in the regular estimation problem, the limiting distribution of $Z^c_T(k)$ is not symmetric around 0. Hence the choice of the intervals may be (more) 'subjective' in the sense that different optimality criteria lead to different optimal intervals.

We distinguish analytical and resampling methods for confidence intervals. In the analytical methods, we can use the obtained asymptotic distributions $F_Z$ with the estimated nuisance parameters – those of the tail. Nuisance parameters enter normalization constants $a_T$ and the parameters that enter the specification of $F_Z$. However, often we can estimate $a_T$ consistently only up to a proportionality constant. In such cases subsampling could fortunately be used to construct confidence intervals.

There are two basic forms of resampling methods – bootstrap and subsampling. The important results in this respect are:

- Under extreme rank condition *bootstrap fails.*
- Under extreme rank condition *subsampling is valid* (with mild additional regularity conditions).

Regression estimators we consider include the no covariate case as a special case, particularly the extreme order statistics. These constitute a well-known example of bootstrap failure, due to Bickel and Friedman (See [62]). On the other hand, subsampling consistency theorems 2.2.1 and 3.2.1 of Politis and Romano[62], given their regularity conditions, combined with the existence of the limit law in Theorem 3 and 5 validate subsampling schemes for rich scenarios. This includes independent as well as dependent data.

We describe here the form of subsampling of [62] suited for time series. Let $W_t = (Y_t, X_t)$. Index by $j$ all subsamples $\{W_i, ...W_{i+b-1}\}$ of the whole sample $\{W_t\}$. There are $N_{b,T} = T - b + 1$ such subsamples. For independent data, preserving order is not necessary, so one can consider all unordered subsamples of length $b$, so that $N_{b,T} = T!/[(T-b)!b!]$.

For each of the subsamples, indexed as $j \in \{1, ...N_{b,T}\}$ we compute the estimator $\hat{\beta}_j(\frac{k}{b})$, and then center it around $\hat{\beta}(\frac{k}{b})$, the estimate based on the full sample:

$$Z^c_{j,T} = a_b \left( \hat{\beta}_j(\frac{k}{b}) - \hat{\beta}(\frac{k}{b}) \right).$$

Law $F_Z$ is then estimated as

$$L_{b,T}(z) = \frac{1}{N_{b,T}} \sum_{j=1}^{N_{b,T}} 1\Big( Z^c_{j,T} \leq z \Big).$$

Rewrite it as :

$$L_{b,T}(z) = \frac{1}{N_{b,T}} \sum_{j=1}^{N_{b,T}} 1\Big( a_b \left( \hat{\beta}_j(\frac{k}{b}) - \beta(\frac{k}{b}) \right) - a_b \left( \hat{\beta}(\frac{k}{b}) - \beta(\frac{k}{b}) \right) \leq z \Big).$$

Intuitively, if $a_b(\hat{\beta}(\frac{k}{b}) - \beta(\frac{k}{b})) \xrightarrow{p} 0$, then the estimator should work by a law of large numbers (for U-statistics). This is the essence of the argument in Politis et al.[62]. Indeed, the *proofs* of the mentioned theorems in [62] require

- $b/T \to 0$ and $b \to \infty$,
- estimated centering can be replaced by the truth consistently (that is , here $a_b(\hat{\beta}(\frac{k}{b}) - \beta(\frac{k}{b})) \xrightarrow{p} 0)$.

Then subsampling is consistent, which means that $L_{T,b}(z) \xrightarrow{p} F_Z(z)$ at continuity points $z$ . This holds uniformly if $F_Z$ is continuous (by Polya's theorem). As a result $L_{T,b}^{-1}(p) \xrightarrow{p} F_Z^{-1}(p)$ at all continuity points $p$ (and uniformly in compact subsets of $(0,1)$ if $F_Z$ is continuous) (cf. [62]). In our case, $F_Z$ is indeed continuous (see Remark on density in Chapter 2).

The key condition that may fail is the second condition. Surprisingly, it does not fail even in the "inconsistency" case when $a_T \to 0$. Indeed, since $\frac{k}{b} \times T \to \infty$, $\hat{\beta}(\frac{k}{b})$ is the sequence of estimators under intermediate ranks, hence under regularity conditions of Theorems 3 and 5:

$$a_b\left(\hat{\beta}(\frac{k}{b}) - \beta(\frac{k}{b})\right) = O_p\left[a_b \frac{F_u^{-1}(m\frac{k}{b}) - F_u^{-1}(\frac{k}{b})}{\sqrt{\frac{k}{b}T}}\right] \text{ and } 0 < m \neq 1$$

but

$$a_b(F_u^{-1}(m\frac{k}{b}) - F_u^{-1}(\frac{k}{b})) \to K$$

by Remark A.4.2. Hence

$$a_b\left(\hat{\beta}(\frac{k}{b}) - \beta(\frac{k}{b})\right) = O_p(\sqrt{b/T}) = o_p(1)$$

Therefore, subsampling is consistent under the regularity conditions stated. Formal statement now follows:

***Corollary 9.1 (Subsampling for Extreme Ranks)***
*Let the data sequence be stationary and strongly mixing. Suppose conditions of Theorem 3 and 5 hold. If $b/T \to 0$ and $b \to \infty$, then subsampling is consistent in the sense defined above, as $\tau T \to k, T \to \infty$.*

The corrolary is also valid if scaling $a_b$ is replaced by a consistent extimator (up to a proportionality constant). Estimation of $a_b$ can be done either directly or by subsampling itself. See [62] and [7] for useful results on the latter. In our empirical work[20], we make a very practical assumption of the exactly algebraic tails and estimate $a_b$ by $b^{-\hat{\xi}}$. Under such condition the estimator of the previous section $\hat{\xi} \xrightarrow{p} \xi$ with an algebraic rate in $T$ (if $\tau T \to$ algebraically fast), so that $b^{-\hat{\xi}}/b^{-\xi} \xrightarrow{p} 1$.

### 5.3.2 Inference under Intermediate Ranks

The plausible feature of the intermediate rank asymptotics is that: (a) the limit law is normal, hence the usual considerations apply in constructing confidence intervals, and (b) normalization constants are easily estimable, as we have demonstrated. Estimation of nuisance parameters $\xi$ and $\mathbf{c}$ has already been discussed. Alternatively, resampling schemes may be used.

Subsampling[1] is *valid* here, and its validation is mainly due to the existence of the limit law (by Theorems 2.2.1, 3.2.1 in Romano and Politis[62], for independent and general dependent settings respectively). As in the previous section, validity of subsampling stems, using the same argument as in the previous section, from noting that (recall that $\tau$ depends on the sample size):

$$a_b\left(\hat{\beta}(\tau_b)-\beta(\tau_b)\right)=a_b\ O_p\Big[\frac{F_u^{-1}(m\tau_b)-F_u^{-1}(\tau_b)}{\sqrt{\tau_b T}}\Big]=O_p\Big[\frac{\sqrt{\tau_b b}}{\sqrt{\tau_b T}}\Big]=o_p(1).$$

***Corollary 9.2 (Subsampling for Intermediate Ranks)***
*Let the data sequence be stationary and strongly mixing. Suppose conditions of Theorem 5 hold. If $b/T \to 0$ and $b \to \infty$, then subsampling is consistent in the sense as defined above, as $\tau T \to \infty, \tau \searrow 0$.*

Implementation of subsampling requires consitent estimation of $a_b$ (the scaling constant under intermediate rank). That is, we should have $\hat{a}_b/a_b \xrightarrow{p} 1$. This is easy given the consistency property of the spacing of the intermediate rank regression quantiles, stated in the Theorem 5. Implementation also requires explicit statement of the rate at which $\tau \searrow 0$.

Properties of bootstrap for indermediate rank statistics are yet unknown. However, given the asymptotic linear representations and limiting normality, one may conjecture validity.

## 5.4 Monte-Carlo Work

In this section we exhibit the essence of Monte Carlo work contained in Appendix C and [19]. This Monte-Carlo strongly supports the theory developed in the previous and current chapters.

[1]By this we mean the schemes in Chapter 2 and 3 in Politis and Romano[62].

### 5.4.1 Setup and Details

The Monte-Carlo considers a simple benchmark model simple benchmark model: $Y_t = X_t'1 + U_t$, $U_t$ are independent of $X_t$; both $\{Y_t, X_t\}$ are i.i.d. Particularly, $X_1 \sim Beta(3,3)$ has a symmetric distribution with compact support $[0,1]$ (in the center, it is close to normal), $X_2 \sim Beta(1.5,6)$ has a continuous skewed distribution. $X_2$ is independent of $X_1$. $X_3$ and $X_4$ are categorical "dummy" variables taking values 1 or 0. $P(X_1 = 1) = 1/3$, $P(X_4 = 1) = 1/3$, $P(X_4 = 1, X_3 = 1) = 0$, $P(X_4 = 0, X_3 = 0) = 1/3$. The design is thus representative of regressors in economic applications: symmetrically and asymmetrically distributed and categorical variables. We considered the three types of tail behavior for the disturbance term $U_t$: type 1 tails (thin tails), type 2 tails (heavy tails), and type 3 tails (bounded support). To represent every type we used standard normal, Cauchy, and uniform distributions.

The sample sizes involved in Monte-Carlo were 50 and 100 (small), 200 and 500 (moderate), and 1000 (large). We selected $k = 1, 5, 10$ as small ranks, and $k = 15$ and 40 as intermediate ranks. Obviously, such division is rather arbitrary, albeit substantiated by intuition.

*Small Ranks:* $k \leq 10$

1. As the sample size increases, our asymptotic theory works as predicted. Small sample discrepancies in approximation disappear as the asymptotic limits are taken.

2. Small sample performance of the extreme rank approximations is *quite remarkable.* For samples as small as 50, the approximations have a high quality, especially for type 2 and type 3 tails. It is not so for type 1 tails, where small sample performance is acceptable, but not as good. The approximation for type 1 tails are accurate for large sample. This agrees with the observations made for non-regression problems [31].

3. *Extreme rank* distributions significantly outperform the central or intermediate rank distributions. They are able to accurately *match the significant kurtosis and skewness* features of the true distributions. The normal (intermediate and central) approximations can not do this in principle.

4. *There is no uniform ranking of the confidence intervals.* Central and intermediate c.i. can be wider or narrower than the c.i. by extreme ranks. This contradicts the intuition derived from the rates of convergence. *This is of value, since extreme ranks may often suggest tighter and better confidence intervals about the unknown parameters.*

5. The intermediate rank and central rank approximations are close to each other as

expected.

*Intermediate Ranks:* $15 \leq k \leq 40$

1. Intermediate and central approximations are good for small samples. For large samples and for large $k$, the accuracy is higher.

2. The intermediate rank and central rank approximations are close to each other as expected. Differences are *very small* as $k$ increases and $k/T$ decreases. *Value of the former* approximation is realized only when nuisance parameters must to be estimated.

3. Surprisingly, the extreme rank approximation is able to beat the normal (intermediate and central) methods for $k = 15$. It is able to *match the pronounced but not very significant kurtosis and skewness* features of the true distributions.

The *fundamental problem* of inference methods based on central asymptotics is the inability to estimate the nuisance parameters well. The most relevant nuisance parameter is the density functions of the disturbance term evaluated at the near-extreme quantiles. The central methods involve estimating the density function *nonparametrically*, which require vast amounts of data to work well. But near-extremes are inherently rare. One may get better estimates by exploiting the extreme value restrictions on the tails. This is exactly what the intermediate rank theory does. Using the tail restrictions, intermediate rank theory states that variance depends on tail index and the tail heterogeneity index. Essentially, using this characterization, one can parametrically estimate this variance.

We thus compared the performance of two inference methods in estimating the true variance of the quantile regression coefficients. We used three criteria to judge performance:

(i) Mean Squared Error,

(ii) Mean Absolute Deviation,

(iii) Median Absolute deviation.

These are defined in the appendix C formally.

We find that on all counts the intermediate methods perform

- strongly better by criteria (i) and (ii),

- weakly better by criterion (iii).

Poor performance of central methods by criterion (i) is indicative of their *strong un-robustness*. (i) strongly weights the outlying estimates – those that are too high or too low. Criterion (ii) does so less strongly. Criterion(iii) measure disregards the un-robustness most strongly. By all three criteria, except for few cases with criterion (iii), the central method is well outperformed.

### 5.4.2 Main Conclusion of Monte-Carlo

We examined whether the theory is basically correct. Our asymptotic approximations have a good performance in small samples and nearly-perfect in large samples. This is an important conclusion, since the probability of mistakes in our proofs is nonnegative.

More substantively, we found that for all tail types and regressors:

*when nuisance parameters are known*

- extreme rank theory offers significantly better approximations than the normal theories (the intermediate and the central) for low ranks ($k \leq 15$); its performance is accurate for small samples and rather precise for large samples;
- intermediate and central rank methods perform accurately in moderate and large samples for not very low ranks ( $k \geq 20$); these approximations are always close to each other, but the value of the former is higher when the nuisance parameters must be estimated;

*when nuisance parameters are unknown*

- centrality based methods (standard modules implemented in S-plus) perform very poorly compared to the extremal methods. This is because central methods attempt to nonparametrically estimate the density functions of $U_t$ evaluated at the near-extreme quantiles. Thus data rarity creates the problems. On the other hand, the extremal methods require only the tail index and tail heterogeneity index estimates that can be obtained parametrically;
- the estimators of tail index and other relevant nuisance parameters, proposed in the current chapter, enjoy good performance in small and large samples.

## 5.5 Conclusion

The aim of this chapter was to provide inference tools to apply our theory to data. We built a theory of inference for the *conditional extremal quantiles*. We introduced the estimators of the tail index and the tail heterogeneity index in the linear quantile regression model. Our estimators are based on the spacings of the sample extremal regression quantiles. We offered methods for constructing confidence intervals.

We also demonstrated through Monte-Carlo work that the developed inference and distribution theory can be valuable in practice. The offered here framework has been used in empirical applications ([20]).

# Conclusion

The present work provided the theoretical framework for studying the conditional extremal response – the near-extreme and extreme conditional quantile functions. In particular:

We exhibited economic models in which the extremal regression quantiles are vital – models of risk, $(S, s)$ adjustment, job search, auctions, sample selection, and many others.

We developed the asymptotic theory of the extremal regression quantiles: limit laws and an inference framework – these results are new.

Particularly, under extremal ranks, we showed that the extremal regression quantile statistic converges in distribution to the argmin of an integral of a point process involving the regressors. Under intermediate ranks, the limit distribution was demonstrated to be normal but the rates of convergence were shown to be non-standard.

These essential results then naturally affirmed various inference techniques and extensions – confidence interval construction and hypothesis testing; importantly, the validation or invalidation of the main resampling methods; characterization of the extremal regression quantile coefficient process behavior in general functional spaces; test statistics as functionals of such processes, and many more.

Significantly, we demonstrated the relevance of this developed theory in the empirical work.[20]

# Appendix A

# Appendix for Chapters 2 and 3

## Appendix

Appendix offers a more general (but more technical) treatment than in the main text, and it provides the proofs of the main results in the form of corollaries. It is organized as follows:

Section A.1 gives useful background definitions on weak convergence of point processes and other useful theorems. In part A.2 the asymptotic distribution theory is developed under extreme rank conditions. It builds general conditions under which the weak convergence of the regression quantile statistics is shown to be determined by the convergence of certain point process. In part A.3 the asymptotic distribution theory is developed under the intermediate rank conditions. Convergence is shown to rely only on suitable CLT, LLN, and analytical properties. Part A.4 gives the essential properties of the model of congenial tail heterogeneity, and part A.5 provides sufficient conditions for the non-degeneracy design conditions made in part A.2.

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 $(\Omega, \mathcal{F}, P)$.

## A.1 Background Definitions and Lemmas

### A.1.1 Weak Convergence of Point Processes

The following definitions may be found in [71] and [56]. They are collected here for readers' convenience.

**Definition A.1.1 (Point Process/ Random Measure)** Let $E$ be a locally compact topological space with a countable basis. Define $\mathcal{E}$ to be the Borel $\sigma$-algebra of subsets of $E$. A *point measure* (p.m.) $N$ on $(E, \varepsilon)$ is a measure of the following form: for $\{x_i, i \geq 1\}$, a countable collection of points (called points of $N$),

$$N(\cdot) \equiv \sum_i 1(x_i \in \cdot).$$

If $N(K) < \infty$, for any $K \subset E$ compact, then $N$ is Radon. A p.m. is simple if $N(x) \leq 1 \quad \forall x \in E$, and is compound otherwise. Let $M_p(E)$ be the collection of all Radon point measures. Similarly, define $M_+(E)$ to be the collection of all Radon nonnegative measures. Clearly $M_p(E) \subset M_+(E)$. Define $\mathcal{M}_p(E)$ and $\mathcal{M}_+(E)$ to be the $\sigma$-algebras generated by the open sets of vague topology on $M_p(E)$ and $M_+(E)$ (see below). A *point process* on $E$ is a measurable map $\mathbf{N}$: $(\Omega, \mathcal{F}, P) \to (M_p(E), \mathcal{M}_p(E))$. A *random measure* on $E$ is a map $\mathbf{N}$: $(\Omega, \mathcal{F}, P) \to (M_+(E), \mathcal{M}_+(E))$.

**Definition A.1.2 (Convergence of Point Processes & Random Measures.)** Sequence $\{N_n\} \subset M_p(E)$ (or $M_+(E)$) converges vaguely to $N$, if $\int f dN_n \to \int f dN$ for all functions $f \in C_K(E)$ (continuous and vanishing outside a compact set) (cf. Leadbetter et..[56]). Vague convergence induces *vague topology* on $M_p(E), M_+(E)$. Topological spaces $M_p(E)$, $M_+(E)$ are metrizable as complete separable metric spaces. Note, $M_p(E)$ is vaguely closed in $M_+(E)$, meaning that if $\{N_n\} \subset M_p(E)$ vaguely converges to $N$, then $N \in M_p(E)$. Given metrizability, the concept of *weak convergence* in $M_p(E)$ (or $M_+(E)$) is the same as for any metric space (cf. [63]): $\mathbf{N}_n \Rightarrow \mathbf{N}$ if $\mathbb{E}h(\mathbf{N}_n) \to \mathbb{E}h(\mathbf{N})$ for all continuous and bounded functions $h$ mapping $M_p(E)$ (or $M_+(E)$) to $\mathbb{R}$.

**Definition A.1.3 (Poisson Point Process or Random Measure (PRM))** Point process $\mathbf{N}$ is a PRM with *mean intensity measure* $m$ (defined on $(E, \mathcal{E})$), if (a) for any $F \in \mathcal{E}$, and any non-negative integer $k$

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

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

### A.1.2 Convex Lower-Semi-Continuous Objectives

The following is a very useful slightly rephrased theorem in Knight[48](Theorem 5) ( Knight attributes the result to C.J.Geyer). This theorem is a very general version of "convexity arguments," first used by Jureckova[45], Davis etal.[25], Pollard[65], among others. It allows for rather general discontinuities. This result is elegantly embedded by Knight[48] into the framework of *weak epi-convergence of lower-semi-continuous objective functions*, that gives a very general way to obtain the weak convergence of argmins. Weak epi-convergence is equivalent to weak finite-dimensional convergence under the condition of *stochastic equi-lower-semi-continuity (s.e-sc.)* and the sequence of agrmins being stochastically bounded. In case of convexity and other properties listed in the theorem, one has s.e.-sc. Related literature on epi-convergence is [74], [76], [27] etc.

***Lemma A.1.1 ( Knight[48], p.12 )***

*Suppose $\{Q_T\}$ is a sequence of lower-semi-continuous (lsc) convex 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 $Q_\infty$ on $\mathcal{D}$ where $Q_\infty$ is lsc convex and finite on an open non-empty set a.s., then $\underset{z\in\mathbb{R}^d}{\text{argmin}}\, Q_T(z) \xrightarrow{d} \underset{z\in\mathbb{R}^d}{\text{argmin}}\, Q_\infty(z)$, provided latter is uniquely defined a.s. in $\mathbb{R}^d$.*

## A.2 Asymptotics under Extreme Ranks

This section derives stochastic limits under extreme rank condition, that is when

$$\tau \searrow 0, \ \tau T \to k, \ T \to \infty.$$

To simplify notational complexity of statements, we always assume that sequence $\{\tau\}$ is s.t.

$$\tau = \frac{k}{T}.$$

### A.2.1 Preliminary Analysis and Basic Assumptions

First, we formulate the fundamental regularity conditions, called Basic Conditions (BC). The BC are meant to be general and simple. Indeed, BC require the key point process, denoted as $\widehat{\mathbf{N}}$, to converge weakly to a well-behaved process $\mathbf{N}$. Process $\widehat{\mathbf{N}}$ characterizes arrival of the conditional extremal events. Such assumption is made a regularity condition since our results *incorporate future* weak convergence results for point processes. This generality is obtained, because the statistics we study is a (as shown) continuous map of the key point process $\widehat{\mathbf{N}}$. This uncovers broad possibilities for establishing the stochastic limit results under general conditions, inclusive of heterogeneous and dependent data. In each such case, we only need to show that BC holds, relying upon the well established theory of point processes (cf.[46], [56], [72]). Indeed, section A.2.4 and A.5 provide some general sufficient conditions.

**Preliminaries.** The key point process mentioned is of the form

$$\widehat{\mathbf{N}}(\cdot) \equiv \sum_{t=1}^{T} 1\{a_T(U_t - b_T), X_t) \in \cdot\}. \tag{A.2.1}$$

where $U_t$ is an error term, suitably defined in terms of $Y_t$ and $X_t$, and $(a_T, b_T)$ are some scaling constants. We construct three basic forms of **mis-en-scenes** that define (1) the state spaces $E$ on which the key point processes $\widehat{\mathbf{N}}$ live, (2) the type of the error terms $U_t$, (3) restrictions on the normalization constants $a_T, b_T$, and (4) the forms of rescaled estimators:

| | | |
|---|---|---|
| **1**. $E_1 \equiv [-\infty, \infty) \times \mathbf{X}$, $U_t \equiv Y_t - X_t'\beta(q)$, $\hat{Z}_T \equiv a_T\big(\hat{\beta}(\tau) - \beta(q)$ $-b_T\mathbf{e}_1\big), a_T > 0;$ | **2**. $E_2 \equiv \big[[-\infty, \infty] \setminus \{0\}\big]$ $\times\mathbf{X}$, $U_t \equiv Y_t - X_t'\beta(\frac{1}{2})$, $\hat{Z}_T \equiv a_T\big(\hat{\beta}(\tau) - \beta(\frac{1}{2})\big)$, $b_T = 0, \ a_T > 0.$ | **3**. $E_3 \equiv [0, \infty) \times \mathbf{X}$, $U_t \equiv Y_t - X_t'\beta(0)$, $\hat{Z}_T \equiv a_T\big(\hat{\beta}(\tau) - \beta(0)\big)$, $b_T = 0, \ a_T > 0.$ |

(A.2.2)

where $q = 0$ in finite support case and $q = \frac{1}{2}$ in the infinite support case; $\mathbf{e}_1 = (1, 0, ..., 0)'$; $\mathbf{X}$ is a compact subset of $\mathbb{R}^d$ that will be assumed to contain the support of $X_t$ $\forall t$. Note that $\widehat{\mathbf{N}}, \mathbf{N}$, scaling constants etc. depend on the type of mis-en-scene ( for simplicity of notation, we do not index them by type).

**Remark A.2.1 (Compactification)** Note that *correct* choice of $E_1$, $E_2$, $E_3$ and the corresponding *topology* is very *important* and *simplifies the proofs* considerably. It is assumed that topology on $E_2$ and $E_1$ is generated by 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 $a < \infty$.)

**Remark A.2.2** Under a particular set of dependence and heterogeneity assumptions of section A.2.4, mis-en-scene 1 suits the cases when $F_{U_t}(\cdot|X_t)$ has type 1 tails (finite and infinite support cases), mis-en-scene 2 – type 2 tails (infinite support), mis-en-scene 3 – type 3 tails.

**Remark A.2.3** The specific forms of $(a_T, b_T)$ have already been discussed in the main text for the benchmark tail-homogeneous i.i.d. case. Section A.2.4 defines them under conditions of dependence (strong mixing) and tail heterogeneity of $F_{U_t}(\cdot|X_t)$ in the covariate $X_t$. Conceivably, these constants may be different under other sets of conditions. Theorem 10 requires only existence of $(a_T, b_T)$ s.t. $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$.

It will be assumed that $\widehat{\mathbf{N}}$ converges weakly to the process $\mathbf{N}$ in $M_p(E)$ (for $E$ equal to one of $E_j$ defined in (A.2.2)) of the form:

$$\mathbf{N}(\cdot) = \sum_{i \geq 1} 1\{(J_i, \mathcal{X}_i) \in \cdot\}, \tag{A.2.3}$$

where $\{J_i, \mathcal{X}_i\}$ are *random vectors* taking values in $E$ a.s.

To overcome certain nuisance problem with non-uniqueness (to be described) define the *perturbed program* and the solution $\hat{\beta}(\tau)$ as:

$$\hat{\beta}(\tau) = \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_{\tilde{\tau}}\big(Y_t - X_t'\beta\big), \quad \text{where } \tilde{\tau} \equiv \frac{\mathbf{k}}{T}, \ \mathbf{k}(k) \equiv k + \nu_T, \tag{A.2.4}$$

where $\nu_T \geq 0$ are small i.i.d. uniformly and independently of $\{Y_t, X_t\}$ distributed shocks with support $\delta_\nu \geq 0$, defined on $(\Omega, \mathcal{F}, P)$. This perturbation device in (A.2.4) eliminates non-uniqueness a.s. in the limit program (as we'll discuss). In cases when at least one covariate is continuously dstributed, we may set $\delta_v = 0$ (no perturbation).

Given such definitions, we state the program which the rescaled statistic $Z_T$ is *defined* to solve:

$$Z_T = \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ Q_T(z, \mathbf{k}) \equiv \sum_{t=1}^{T} \rho_{\widetilde{\tau}}\big(a_T(U_t - b_T) - X_t' z\big) \right\} \tag{A.2.5}$$

(which can be seen to correspond to (A.2.4) by writing $z$ as corresponding to the mis-en-scenes: $z \equiv a_T(\beta - \beta(q) - b_T \mathbf{e}_1)$ or $z \equiv a_T(\beta - \beta(\frac{1}{2}))$ or $z \equiv a_T(\beta - \beta(0))$.

The conjectured weak limit, as we claim, will solve the problem (for $E$ equal to one of the space types $E_i$):

$$Z = \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ Q_\infty(z, \mathbf{k}) \equiv -\mathbf{k}\mu_X' z + \int_E l(u, x'z) d\mathbf{N}(j, \mathcal{X}) \right\} \tag{A.2.6}$$

where $l(u, v) \equiv 1(u \leq v)(v - u)$, $\mu_X =$ plim $\bar{X}$. Since $\mathbf{N} \in M_p(E_i)$, the objective will always be well defined and finite on some open non-empty subset of $\mathbb{R}^d$, under conditions to be stated. Given $\omega \in \Omega$, denote by $\mathcal{K}_b(w)$ the set of points $k_b \in \mathcal{K} = [K_1, K_2]$ (s.t. $k$ is in interior of $\mathcal{K}$) s.t. $\min_{z \in \mathbb{R}^d} Q_\infty(z, k_b)$ does not have a unique solution (called *break-points*) in $\mathbb{R}^d$. For any $\omega$, such break-points exist due to piecewise-linear nature of the objective (see gradient conditions in the main text). We assume **essential uniqueness**: that is, (i) $k \notin \mathcal{K}_b$ w.p. 1 (then no perturbation is needed– set $\delta_v = 0$, i.e. $\mathbf{k} = k$) or (ii) Lebesgue measure of $\mathcal{K}_b$ is zero w.p.1 (then perturbation eliminates non-uniqueness a.s.). We'll justify this assumption.

Above discussion motivated the following set of simple regularity conditions.

***Assumption 3 (BASIC CONDITIONS)***

**BC(1)** *(Convergence of the Key Point Process) There exists a sequence of constants $\{a_T, b_T\}$ and $E_i$ of one of the forms stated in (A.2.2) s.t. point process $\widehat{\mathbf{N}}$ defined in (A.2.1) converges weakly in $M_p(E_i)$ to a process $\mathbf{N}$:*

$$\widehat{\mathbf{N}} \Rightarrow \mathbf{N}.$$

*(For space type $E_2$ we only require $\widehat{\mathbf{N}}(\cdot \cap E_2') \Rightarrow \mathbf{N}(\cdot \cap E_2')$ in $M_+(E_2')$, where $E_2'$ $= [-\infty, 0) \times \mathbf{X}$). The points of $\mathbf{N}$ are assumed to be $\mathbb{R}^{d+1}$-valued a.s.*

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

**BC(3)** *Design Conditions (Essential uniqueness and a No-crossing Property).*

*(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.*

**Remark A.2.4** (1) As discussed, convergence of $\widehat{\mathbf{N}}$ to $\mathbf{N}$ is the key, since all weak limits are shown to be continuous functionals of the random measure $\mathbf{N}$. Note that points of $\mathbf{N}$ are finite a.s. BC(2) requires of $\{X_t\}$ and $\{\mathcal{X}_t\}$ to have basic stability properties. Limit empirical distribution function exists e.g. under stationarity and ergodicity or whenever a LLN holds. Compactness condition is required for establishing the continuity of mapping $\mathbf{N}$ to the rescaled statistic. Relaxing this condition may change the limit results depending on the tail behavior of $X_t$. (3) Condition BC(3)(1), given BC(2), is plausible (see section A.5). Potentially, BC(3)(1) covers the limits $\mathbf{N}$ that are compound Poisson Processes that have clusters at its points (see [56]). (4) Condition BC(3)(2) says that for space type $E_2$, the population regression median is below or equal to low rank sample regression quantiles in $\mathbf{X}$ with arbitrarily small probability as $T \to \infty$. This is a plausible condition, and it is easy to see it holds in the settings of the main text.

### A.2.2 Asymptotic Distributions under Extreme Ranks

The following is **a main result**.

***Theorem 10 (Weak Convergence of the Quantile Regression Statistics)***
*Under conditions BC(1)-(3) (for appropriate space type E) as $\tau T \to k, T \to \infty$:*

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

*where $l(u,v) = 1(u \le v)(v-u)$ and $\mathbf{N}$ is defined in (A.2.3).*

Proof appears after the following corollary. This result could be rephrased in terms of the centered at $\beta(\tau)$ statistics:

$$Z_T^c \equiv a_T\big(\hat{\beta}(\tau) - \beta(\tau)\big)$$

Assume in addition the following basic condition:

***Assumption 4***

**BC(4)** *For corresponding mis-en-scene, as $\tau T \to k, T \to \infty$, one of the following is true (for $\mathbf{c}(k) \in \mathbb{R}^d$): $a_T(\beta(\tau) - \beta(q) - b_T\mathbf{e}_1) \to \mathbf{c}(k), a_T(\beta(\tau) - \beta(\frac{1}{2})) \to \mathbf{c}(k)$, or $a_T(\beta(\tau) - \beta(0)) \to \mathbf{c}(k)$.*

***Corollary 10.1 (Convergence of Centered Estimates)***
*Given conditions BC(1)-BC(4), the following weak limits are true (for appropriate mis-en-scene):*

$$Z_T^c \xrightarrow{d} Z^c \equiv Z - \mathbf{c}(k), \text{ as } \tau T \to k, T \to \infty.$$

Examples of $\mathbf{c}(k)$ were given in the main text.

**Proof of Theorem 10**: This proof relies on the theory of weak convergence of point processes and on the "convexity" theorem 5 in Knight[48], restated in appendix as lemma A.1.1.

(**1**) First, stabilize objective function $Q_T(z, \mathbf{k})$ in (A.2.5) by subtracting a term that does not affect optimization in (A.2.5) :

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

where $l_\delta(u, v) \equiv 1(u \le v)(v - u) - 1(u \le -\delta)(-\delta - u)$, for $\delta > 0$. Such renormalization also makes the proof particularly simple[1], since $\widetilde{Q}_T$ can be rewritten as a (to be shown) continuous functional of $\widehat{\mathbf{N}}$:

$$\widetilde{Q}_T(z, \mathbf{k}) = -\mathbf{k}\bar{X}'z + \int_{E_i} l_\delta(j, x'z) d\widehat{\mathbf{N}}(j, x). \tag{A.2.7}$$

Part (2) (verifying continuity) of the proof shows that fidi distributional limit of $\widetilde{Q}_T$ is the convex function:

$$\widetilde{Q}_\infty(z, \mathbf{k}) = -\mathbf{k}\mu_X'z + \int_{E_i} l_\delta(j, x'z) d\mathbf{N}(j, x). \tag{A.2.8}$$

$\widetilde{Q}_T(z, \mathbf{k})$ is convex $\forall T$, since it is a sum of convex functions in $z$. $Q_T(\cdot, \mathbf{k})$ is continuous, hence l.s.c. $\forall T$. By theorem 5 in [48] , we have that[2]

$$Z_T \xrightarrow{d} Z \equiv \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \widetilde{Q}_\infty(z, \mathbf{k}) \tag{A.2.9}$$

provided (i) there is a non-empty open set $\mathcal{Z}_0$ s.t. $\widetilde{Q}_\infty(z, \mathbf{k})$ is finite a.s. $\forall z \in \mathcal{Z}_0$, and (ii) $Z$ exists and is a.s. unique. (i) is shown in part 2 of this proof. (ii) simply follows

[1] Proving the result with $l(u, v) = 1(u \le v)(v - u)$ in place of $l_\delta(u, v)$ also works, with the difference being that map $N \mapsto \int_E l dN$ is a.s. continuous under BC whereas $N \mapsto \int_E l_\delta dN$ is continuous.

[2] See the remark after the proof that deals with case of $E_2'$ separately.

from BC(3)(1), since $Z =$ argmin $_z Q_\infty(z, \mathbf{k})$, for $Q_\infty$ defined in equation (A.2.6). Indeed, $Q_\infty(z, \mathbf{k})$ differs from $\widetilde{Q}_\infty(z, \mathbf{k})$ only by $\Delta = \int_E 1(j \leq -\delta)(-\delta - j)d\mathbf{N}(j, x) = \sum_{i:J_i \leq -\delta}(-\delta - J_i)$, which is independent of $z$ and $\Delta \neq \infty$ or $-\infty$ a.s. [3]

(**2**) Here we verify that (**i**) $\widetilde{Q}_\infty(\cdot, \mathbf{k})$ is indeed a fidi distributional limit, (**ii**) there is an open non-empty set $\mathcal{Z}_0$ s.t. $\widetilde{Q}_\infty(z, \mathbf{k})$ is finite a.s. for $z \in \mathcal{Z}_0$.

$\widetilde{Q}_\infty(\cdot, \mathbf{k})$ is a distributional fidi limit of $\{\widetilde{Q}_T(\cdot, \mathbf{k})\}$ iff for any finite collection $(z_j, j \leq l)$

$$\left(\widetilde{Q}_T(z_1, \mathbf{k}), ..., \widetilde{Q}_T(z_l, \mathbf{k})\right) \xrightarrow{d} \left(\widetilde{Q}_\infty(z_1, \mathbf{k}), ..., \widetilde{Q}_\infty(z_l, \mathbf{k})\right).$$

Since $\bar{X}' z_j \xrightarrow{p} \mu_X' z_j$ and $\mathbf{k} \xrightarrow{d} \mathbf{k}$, independently of $\{a_T(U_t - b_T), X_t\}$, we need to verify only that for $z_1, ..., z_l$ :

$$\left( \int_{E_i} l_\delta(u, x' z_j) d\widehat{\mathbf{N}}(u, x),\ j \leq l \right) \xrightarrow{d} \left( \int_{E_i} l_\delta(u, x' z_j) d\mathbf{N}(u, x),\ j \leq l \right). \qquad \text{(A.2.10)}$$

(**a**) Take the case of $E_1$ (mis-en-scene 1), then note that 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 compact subset $K$ *in* $E_1$:

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

$K$ is compact in $E_1$ since $\kappa < \infty$ by BC(2). Define map (for $E_i = E_1, E_2', or E_3$)

$$\mathbf{T}_i : \widehat{\mathbf{N}} \mapsto \left( \int_{E_i} l_\delta(u, x' z_j) \ d\widehat{\mathbf{N}}(u, x),\ j \leq l \right)$$

Hence it follows $\widehat{\mathbf{N}} \mapsto \mathbf{T}_1(\widehat{\mathbf{N}})$ is continuous from $M_p(E_1)$ to $\mathbb{R}^l$. Thus $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$ implies $\mathbf{T}_1(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}_1(\mathbf{N})$ by the continuous mapping theorem.

(**b**) Take the case of $E_3$ (mis-en-scene 3). Note that map $(u, x) \mapsto l(u, x' z_j)$ is in $C_K(E_3)$, since it is uniformly continuous on $E_3$ by construction and vanishes outside compact subset $K$ *in* $E_3$:

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

[3] Indeed, in case of $E_3$, $j \geq 0$, hence $\Delta = 0$. In case of $E_1$, $E_2$, note (i) $\mathbf{N}(K) < \infty$ a.s. for all $K$ compact (by definition of $\mathbf{N} \in M_p(E_i)$). $K = [-\infty, -\delta] \times \mathbf{X}$ is indeed a compact subset of $E_2$ or $E_3$ (see remarkA.2.1), $\Rightarrow \#\{i : J_i \leq -\delta\} < \infty$ a.s., (ii) points $\{J_i\}$ of $\mathbf{N}$ are real-valued by $BC(1)$. (i) and (ii) $\Rightarrow \Delta \neq \infty$ or $-\infty$ a.s. by BC(1)

$K$ is compact in $E_3$ since $\kappa < \infty$ by BC(2). Therefore, it follows $\widehat{\mathbf{N}} \mapsto \mathbf{T}_3(\widehat{\mathbf{N}})$ 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 the case of space $E_2$ (mis-en-scene 2). We claim that we need to show the weak fidi convergence (A.2.10) *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. $\to 1$, as $T \to \infty$, by BC(3). This claim is verified in the remark after the proof. 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 compact subset $K$ *in* $E_2'$:

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

$K$ is compact in $E_2'$ since $\kappa < 0$ by BC(2) 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 $\mathbf{T}_2(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}_2(\mathbf{N})$.

We have shown (**i**). In order to show (**ii**), take distinct $(z_1, \dots, z_l, l \geq d+1)$ not on a hyperplane in $\mathbb{R}^d$ ( for $E_2'$, $z_i \in \mathcal{Z}_N$) and take $\mathcal{Z}_0$ as the interior of convex hull of these points. Obviously, then for any $z \in \mathcal{Z}_0$ $(u, x) \mapsto l_\delta(u, x'z) \in C_K(E_i)$, by the given above arguments, which implies $\int_{E_i} l_\delta(u, x'z) d\mathbf{N}(u, x)$ is finite a.s. [Note: $(u, x) \mapsto l_\delta(u, x'z) \in C_K(E_i) \Rightarrow$ (a) $\#\{t : l_\delta(J_t, \mathcal{X}_t'z) \neq 0\}$ is finite a.s. and (b) $l_\delta(u, x'z)$ is bounded on $E_i$, (a) & (b) $\Rightarrow$ the finiteness a.s. ] ■

**Remark A.2.5** Consider the case of space $E_2'$ (mis-en-scene 2). We claimed in the proof that we need to show the weak fidi convergence (A.2.10) for points $\mathcal{Z}_N = \{z : x'z < 0, \forall x \in \mathbf{X}\}$ only, since $Z_T$ and $Z$ necessarily belong to such set w.p. $\to 1$ as $T \to \infty$ by BC(3). The claim is verified by considering the objective functions:

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

where $\phi(A) = 0$, if $A$ is true, $\phi(A) = \infty$ if not. As above, it is not difficult to verify that the objectives are convex and l.s.c as well as finite on an open non-empty set. Hence by the convexity lemma in A.2, $Z_T^\epsilon \equiv$ argmin ${}_z\widetilde{Q}_T(z, \mathbf{k}, \epsilon) \xrightarrow{d} Z^\epsilon \equiv$ argmin ${}_z\widetilde{Q}_\infty(z, \mathbf{k}, \epsilon)$ by fidi convergence demonstrated as in the proof of theorem 1 (part 2 (c), except that here if $z_j$ is s.t. $\phi(\cdot) = \infty$, $\widetilde{Q}_T(z_j, \mathbf{k}, \epsilon) = \infty \to \widetilde{Q}(z_j, \mathbf{k}, \epsilon) = \infty$, trivially), conditionally on the event that $Z^\epsilon$ is uniquely defined. Then we can choose $\epsilon$ small, s.t. the probability that $Z_T^\epsilon$ and

$Z^\epsilon$ differ from $Z_T$ and $Z$, is arbitrarily small as $T \to \infty$ by BC(3)(2). Thus the probability that $Z^\epsilon$ is not uniquely defined is arbitrarily small by BC(3)(1). This shows $Z_T \xrightarrow{d} Z$.

### A.2.3 Joint Convergence in Distribution

Define

$$Z_T(k) \equiv \text{argmin}_{z\in\mathbb{R}^d} Q_T(z, \mathbf{k}(k)) \text{ and } Z(k) \equiv \underset{z\in\mathbb{R}^d}{\text{argmin}}\, Q_\infty(z, \mathbf{k}) \tag{A.2.11}$$

For a finite collection $(k_j,\ j \le l)$, what is the limit joint distribution of $\big(Z_T(k_j),\ j \le l\ \big)$ ?

***Theorem 11 (Joint Distribution)***
*Under conditions BC(1)-BC(3)*

$$\mathbf{Z}_T \equiv \big(Z_T(k_j)', j \le l\big)' \xrightarrow{d} \mathbf{Z} \equiv \big(Z(k_j)',\ j \le l\ \big)'.$$

**Proof of Theorem 11**

$$\mathbf{Z}_T \in \underset{z=(z_1,\dots,z_l)\in\mathbb{R}^{d\times l}}{\text{argmin}} \widetilde{Q}_T(z_1, \mathbf{k}_1) + \dots + \widetilde{Q}_T(z_l, \mathbf{k}_l).$$

Since the new objective function is a sum of objective functions in the previous theorem, it retains the properties of elements being summed, since the new objective function is separable in the components of $\mathbf{Z}_T$. Therefore the argument of Theorem 1 applies. ■

### A.2.4 Types of Limit Point Processes N under Heterogeneity and Dependence

This section consists of three lemmas. Lemma A.2.1 (due to Resnick[71]) shows two essential conditions for a convergence to a simple point process $\mathbf{N}$ in $M_p(E)$. Lemma A.2.2 (due to Meyer[59]) shows that if data is *weakly dependent*, then only one condition suffices. Lemma A.2.3 gives a *congenial model of tail heterogeneity in covariate* s.t. this condition is true.

*General Results*

In order for the general theory to hold, it is required by BC(1) that $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$. There are many results that provide this, including the following powerful result.

***Lemma A.2.1 (Resnick[71], Prop. 3.22)***
*Suppose* $\mathbf{N}$ *is a simple point process on* $E$ *and* $\mathcal{T}$ *is a basis of relatively compact open sets s.t.* $\mathcal{T}$ *is closed under finite unions and intersections and for* $F \in \mathcal{T}$, $P(\mathbf{N}(\partial F) = 0) = 1$. *If* $\{\widehat{\mathbf{N}}\}$ *is a sequence of point processes on* $E_i$ *and for all* $F \in \mathcal{T}$:

$$\lim_{T \to \infty} P[\widehat{\mathbf{N}}(F) = 0] = P[\mathbf{N}(F) = 0] \tag{A.2.12}$$

$$\lim_{T \to \infty} \mathbb{E}\widehat{\mathbf{N}}(F) = \mathbb{E}\mathbf{N}(F) < \infty \tag{A.2.13}$$

*then* $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ *in* $M_p(E)$.

**Remark A.2.6** In our case, $\mathcal{T}$ consists of finite unions and intersection of bounded open rectangles in $E_1$, $E_2'$, $E_3$. Refer to remark A.2.1 concerning the topology on $E_1$, $E_2'$, $E_3$.

This lemma shows that the key conditions are (A.2.12) and (A.2.13). The following lemma shows that as long as (A.2.13) holds, (A.2.12) obtains if data is a stationary strongly mixing sequence. The following lemma gives the Meyer[59] mixing condition on events $A_t^T(F) \equiv \{w \in \Omega : (a_T(U_t - b_T), X_t) \in F\}$, adopted here to suit our notation.

***Lemma A.2.2 (Poisson Limits for Strong Mixing Data)***
*Suppose that for any* $F \in \mathcal{T}$, *triangular sequence of events* $\left\{\left(A_t^T(F), t \leq T\right), T \geq 1\right\}$ *is stationary and* $\alpha$ *-mixing with the mixing coefficient* $\alpha_T(\cdot)$, *(A.2.13) holds, and there exist sequences of integers* $(p_n, n \geq 1)$, $(q_n, n \geq 1)$ , $(t_n = np_n + q_n, n \geq 1)$: *as* $n \to \infty$, *for some* $r > 0$ *(a)* $n^r \alpha_n(q_n) \to 0$, *(b)* $q_n/p_n \to 0$, *(c)* $p_{n+1}/p_n \to 1$, *and (d) that* $I_{p_n} = \sum_{i=1}^{p_n}(p_n - i)P(A_1^{t_n}(F) \cap A_{i+1}^{t_n}(F)) = o(1/n)$. *Then in* $M_p(E)$, $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$, *a PRM with mean measure* $m : m(F) \equiv \lim_{T \to \infty} E\widehat{\mathbf{N}}(F)$.

**Proof**. At first, for any $F : m(F) > 0$, $\lim_{T \to \infty} P[\widehat{\mathbf{N}}(F) = 0] = P[\mathbf{N}(F) = 0] = e^{-m(F)}$, by Meyer theorem[59]. This also holds if $m(F) = 0$, i.e. $E\widehat{\mathbf{N}}(F) \to 0$, since it implies $P(\mathbf{N}(F) > 0) \to 0$, [since $\widehat{\mathbf{N}}(F)$ is integer-valued]. Conclude by lemma A.2.1 and definition of PRM. ■

**Remark A.2.7** Condition $I_{P_n} = o(1/n)$ prevents clusters or multiple appearance of "rare" events $A_t^T(F)$. Hence it eliminates compound Poisson processes as limits. The Meyer mixing condition is a generalization of a result due to Loynes[57]. Leadbetter et al[56] give various other generalizations ('distributional mixing') that closely suit specific forms of the point processes with no covariates that they consider.

### A Model of Congenial Tail Heterogeneity

The following is the tail condition that allows one to flexibly account for a possible dependence of the tail on the covariate. It was discussed in the main text, and it may be relaxed in many directions to the degree that the key point process $\widehat{\mathbf{N}}$ converges to a well-behaved process $\mathbf{N}$, as required in BC.

***Assumption 5 (Congenial Tail Heterogeneity in Covariate)***
*Assume the following tail equivalence condition, for $U_t$ defined in (A.2.2):*

$$F_{u_t}(z|X_t = x) \sim K(x)F_u(z) \text{ as } z \searrow -\infty \text{ or } z \searrow 0 \tag{A.2.14}$$

*uniformly in $x \in \mathbf{X}$ and $t$, $F_u$ has type 1-3 tails (and tail index $\xi$),*[4] *$K(\cdot)$ is assumed to be a uniformly positive bounded continuous function on $\mathbf{X}$, normalized so that $K(\mu_X) = 1$. W.l.o.g. $\mu_X = (1, 0, ...)$ (if $\mu_X \notin \mathbf{X}$, pick any other reference point $x^r \in \mathbf{X}$, s.t. $K(x^r) = 1$).*

*Linearity*, assumed throughout $[F_{Y_t}(\tau|X_t = x) = x'\beta(\tau)$ w.p.1], *implies* that (see Lemma A.4.1):

$$K(x) = \begin{cases} e^{-x'\mathbf{c}} & \text{for type 1 tails,} \\ (x'\mathbf{c})^{\alpha} & \text{for type 2 tails,} \\ (x'\mathbf{c})^{-\alpha} & \text{for type 3 tails,} \end{cases} \tag{A.2.15}$$

Constant vector $\mathbf{c}$ is s.t. $K(\mu_X) = 1$. In case of tail homogeneity, $\mathbf{c} = \mathbf{0}$ in the type 1 tails, and $\mathbf{c} = (1, 0, ...)' = \mathbf{e}_1'$ for the others. 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 $\mathbf{X}$.

***Theorem 12 (Limit $\mathbf{N}$ under Congenial Heterogeneity )***
*Under BC(2), Assumption 5 and independence or data-dependence condition in lemma A.2.2, for the canonical normalization constants $(a_T, b_T)$ defined in terms of $F_u$ in the text, $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E_i)$, a PRM with mean measure $m(\cdot)$ 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 $\mathbf{N}$ have representation*

$$(J_i, \mathcal{X}_i, i \geq 1) \stackrel{d}{=} (h^{-1}(\Gamma_i/K(\mathcal{X}_i)), \mathcal{X}_i, i \geq 1),$$

[4] See main text for definition.

*where $h^{-1}$ is the inverse of $h$, $\Gamma_i = \mathcal{E}_1 + ... + \mathcal{E}_i$ ($\{\mathcal{E}_i\}$ are i.i.d. standard exponential), and $\{\mathcal{X}_i\}$ are i.i.d. r.v. with law $F_X$, independently distributed from $\{\Gamma_i\}$.*

**Remark A.2.8** Note that in view of (A.2.15), the points of $\mathbf{N}$ have distribution:

$$(J_i, \mathcal{X}_i) \stackrel{d}{=} \begin{cases} (\ln(\Gamma_i) + \mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 1,} \\ (-\Gamma_i^{-1/\alpha}\mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 2,} \\ (\Gamma_i^{1/\alpha}\mathcal{X}_i'\mathbf{c}, \quad \mathcal{X}_i) & \text{for type 3.} \end{cases} \tag{A.2.16}$$

**Remark A.2.9** In view of (A.2.15), lemma A.4.1 computes the centering constants $\mathbf{c}(k)$ needed to express the limit distribution of centered statistics $Z_T^c = a_T(\hat{\beta}(\tau) - \beta(\tau))$:

$$\mathbf{c}(k) = \begin{cases} \ln k\mathbf{e}_1 + \mathbf{c}, & \text{for type 1,} \\ -k^{\frac{-1}{\alpha}}\mathbf{c}, & \text{for type 2,} \\ k^{\frac{1}{\alpha}}\mathbf{c}, & \text{for type 3.} \end{cases} \tag{A.2.17}$$

***Corollary 12.1***

*Combine Theorems 10-12 with section D to have Theorems 1-3 of the main text.*

**Proof of Theorem 12:** Under stationarity, in view of previous lemmas, this reduces to showing that the mean intensity measure $E\widehat{\mathbf{N}}$ converges for all $F$ in $\mathcal{T}$ to $m$, which is by mimicking convergence proofs for point processes with no covariates (e.g. [71] or [56]). Proof for independent (nonstationary) case is similar.

To show the second part, construct a PRM with given intensity measure $m$. At first, define homogeneous PRM $\mathbf{N}_1$ with points $\{\Gamma_i\}$ (defined as above). It has mean measure $m_1(du) = du$ on $[0, \infty)$ (see [71]). Secondly, by Proposition 3.8 in [71], PRM $\mathbf{N}_2$ with points $\{\Gamma_i, \mathcal{X}_i\}$ (as above) has mean measure $m_2(du, dx) = du \times F_X(dx)$ on $[0, \infty) \times \mathbf{X}$, since $\{\mathcal{X}_i\}$ are i.i.d. and are independent of $\{\Gamma_i\}$. Finally, $\mathbf{N}$ with transformed points $\{\mathbf{T}(\Gamma_i, \mathcal{X}_i)\}$, where $\mathbf{T} : (u, x) \mapsto (h^{-1}(u/K(x)), x)$ has the desired mean measure $m(dj, dx) = m_2 \circ \mathbf{T}^{-1}(dj, dx) = K(x)h(dj) \times F_X(dx)$, by Proposition 3.7 in [71]. ■

## A.3 Asymptotics under Intermediate Ranks

### A.3.1 Asymptotic Distributions under Intermediate Ranks

Define the following key variables:

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

where $\mu_t(l,z) = \big(1(Y_t \leq X_t'\beta(\tau l)) - 1(Y_t \leq X_t'\beta(\tau l) + \frac{X_t'z}{a_T(l)})\big)$. Now we can state the basic conditions:

***Assumption 6 (BASIC CONDITIONS)***

**BC*(1)** *(Analytical Property)* $\lim_T E_P G_T(l,z) = z'\frac{1}{2}J(l)z$, *$J(l)$ is an invertible matrix $\forall l > 0$ for properly chosen $a_T(l)$.*

**BC*(2)** *(LLN)* $\Big\|G_T(l,z) - E_P G_T(l,z)\Big\| \xrightarrow{p} 0$ *for any fixed $l$ and $z$, for a sequence of constants $a_T(l)$*

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

**Remark A.3.1** (i) The analytical condition BC*(1) is **most important**. Under *independence* it implies BC*(2), and BC*(3) holds by Lindeberg-Feller. To see this note $G_T(z,l) \equiv \frac{1}{\sqrt{\tau T}} \sum_t r_t(z,l)$, BC*(1) requires $Er_t(z,l) \sim \sqrt{\tau T^{-1}}\frac{1}{2}z'J(l)z$. Then by compactness of $\mathbf{X}$ and "binomiality" of $r_t$:

$$Var(\sum_t r_t(z,l)/\sqrt{\tau T}) = O(Er_t(z,l)/\tau) \to 0.$$

(ii) BC*(1) is *purely* an analytical condition. Next subsection shows how it obtains for the tail-heterogeneous model from the assumed *regular variation* properties of $\partial F_u^{-1}(\tau)/\partial\tau$. (iii)Under dependence, there are mixing options that when used *carefully* imply BC*(1) and BC*(2)(see [24], [29], [35] for relevant reviews). *Carefulness* means that CLTs should require strictly no more than two bounded moments. As a result, Liapunov CLT and most CLTs on $L_2$-mixingales and $L_2$-NED don't apply; but Lindeberg-Feller and $L_1$-mixingale CLTs do. Here, with some adaptation, we can use local CLTs in Bierens[8] and Robinson[73], designed for kernel estimators.

***Theorem 13 (Weak Convergence under Intermediate Ranks)***
*Under conditions BC*(1)- BC*(3), as $\tau T \to \infty, \tau \searrow 0$: $\big(a_T(l_i)(\hat{\beta}(l_i\tau) - \beta(l_i\tau), i \le m\big) \xrightarrow{d} N(0, \Omega)$, where $\Omega_{ij} = J^{-1}(l_i)\mathcal{G}_{ij}J^{-1}(l_j)$.*

**Proof of Theorem 13** Case of $m = 1$ is very similar to the proof in the main text. The proof for $m \ge 2$ is also obviously analogous: The objective function, that $\mathbf{Z}_T \equiv (Z_T(l_i), i \le m)$ (for $Z_T(l) \equiv a_T(l)\big(\hat{\beta}(l\tau) - \beta(l\tau)\big)$) minimizes, can be written as: $Q_T(\mathbf{z} \equiv (z_i, i \le m)) = \sum_{i=1}^{m} Q_T(z_i, l_i)$, where $Q_T(z, l)$ is defined as in the main text. Hence, by BC*(1),(3), the fidi distributional limit is $Q_\infty(\mathbf{z}) = \sum_{i=1}^{m} W_\infty(l_i)'z_i + \frac{1}{2}z_i'J(l)z_i$, where $W_\infty(l_i), i \le m = N(\mathbf{0}, \mathcal{G})$. $Q_\infty(\mathbf{z})$ has unique minimum at $\mathbf{Z} = \big(-J^{-1}(l_i)W_\infty(l_i), i \le m\big)$, and is finite a.s. on an open non-empty set. Both $Q_\infty(\mathbf{z})$ and $Q_T(\mathbf{z})$ are convex. Conclude $\mathbf{Z}_T \xrightarrow{d} \mathbf{Z}$ by conexity theorem (e.g. Theorem 5 in [48]). ■

### A.3.2 Limits Under Tail Heterogeneity (and Serial Dependence)

#### Analytical Condition

This section gives sets of general sufficient conditions for the analytical condition BC*(1). Im view of the previous remark, this suffices to obtain the stated above result under independence. For the congenial tail heterogeneity model stated in section B, assume in addition for any $x \in \mathbf{X}$ (compact), as $\tau \searrow 0$: (for $U_t$ defined as before)

$$\partial F_{u_t}^{-1}(\tau | X_t = x)/\partial\tau \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau \in \mathcal{R}_{-\xi-1} \tag{A.3.18}$$

($\sim$ holds *uniformly* in $x \in \mathbf{X}$ since $\partial F_{u_t}^{-1}(\tau | X_t = x)/\partial\tau$ is linear (convex) in $x$.)

Scaling $a_T(l)$ is defined as in Theorem 5.

***Lemma A.3.1 (Analytical Property)***
*Suppose $F_X$, the limit distribution function of $\{X_t\}$, with support $\mathbf{X}$, is non-degenerate in $\mathbb{R}^d$. In the stated tail heterogeneity model, BC*(1) holds, with $J(l)$ given in Theorem 5.*

**Proof:** We need to show BC*(3) holds. We suppress $l$ in notation ($l = 1$), since computation are identical for all $l$. 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 $f_{u_t}(\cdot|X_t)$:

$$\begin{aligned}
\frac{E\mu_t X_t' z}{\sqrt{\tau T}} &= \frac{1}{\sqrt{\tau T}} E|F_t(F_t^{-1}(\tau)) - F_t(F_t^{-1}(\tau) + \frac{X_t' z}{a_T}))||X_t' z| \\
&= \frac{1}{\sqrt{\tau T}} E|\frac{f_t(F_t^{-1}(\tau)[1 + o(F_u^{-1}(m\tau) - F_u^{-1}(\tau))])}{a_T}|z' X_t X_t' z \\
&\sim \frac{1}{\sqrt{\tau T}} E|\frac{f_t(F_t^{-1}(\tau))}{a_T}|z' X_t X_t' z \\
&= \frac{1}{T} E|\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_t[F_t^{-1}(\tau)])^{-1}}|z' X_t X_t' z \\
&\sim \frac{1}{T} E\frac{1}{\mathcal{H}(X_t)}|\frac{m^{-\xi} - 1}{-\xi}|z' X_t X_t' z, \text{ uniformly in } t.
\end{aligned} \tag{A.3.19}$$

The first equality follows from the definition of $1/a_T = o(F_u^{-1}(m\tau) - F_u^{-1}(\tau))$ and Taylor expansion. The first equivalence $\sim$ follows from the regular variation properties and uniform in $t$ tail-equivalence: $1/f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau \in \mathcal{R}_{-\xi-1}$. Indeed, by the regular variation property, *locally uniformly* in $l$ (i.e. uniformly in $l$ in any compact subset of $(0, \infty)$

$$f_u(F_u^{-1}(l\tau)) \sim l^{\xi+1} f_u(F_u^{-1}(\tau)). \tag{A.3.20}$$

I.e. locally uniformly in $l$

$$f_u(F_u^{-1}(\tau) + [F_u^{-1}(l\tau) - F_u^{-1}(\tau)]) \sim l^{\xi+1} f_u(F_u^{-1}(\tau)). \tag{A.3.21}$$

Letting $l_\tau \to 1$, we have $f_u(F_u^{-1}(\tau) + [F_u^{-1}(l_\tau\tau) - F_u^{-1}(\tau)]) \sim f_u(F_u^{-1}(\tau))$. Hence if one has sequence $v_\tau = o([F^{-1}(m\tau) - F^{-1}(\tau)])$, for any $0 < m \neq 1$, then as $\tau \searrow 0$:

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

Since uniformly in $t$ $1/f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau$ by assumption, reasoning as in (A.3.20) and (A.3.21) gives that uniformly in $t$, $f_t(F_t^{-1}(\tau) + m_\tau) \sim f_t(F_t^{-1}(\tau))$. Apply this to have first equivalence $\sim$ in (A.3.19).

The second equivalence $\sim$ in (A.3.19) follows from(A.3.22)-(A.3.24). At first, $f_t(F_t^{-1}(\tau)) \sim \partial F_u^{-1}(\tau/K(x))/\partial\tau = 1/(K(x) f_u[F^{-1}(\tau/K(x))])$, uniformly in $t$. Hence, uniformly in $X_t = x$:

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(f_t[F_t^{-1}(\tau)])^{-1}} \sim \frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau(K(x) f_u[F_u^{-1}(\tau/K(x))])^{-1}}. \tag{A.3.22}$$

But by Lemma A.4.1(v) and Remark A.4.2(ii), uniformly in $x \in \mathbf{X}$:

$$\frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau/K(x)) - F_u^{-1}(\tau/K(x))} \to 1/\mathcal{H}(x), \tag{A.3.23}$$

where $\mathcal{H}(x) = 1$ if $\xi = 0$, $\mathcal{H}(x) = x'\mathbf{c}$ if $\xi \neq 0$. And by the regular variation property (A.3.20):

$$\begin{aligned} \frac{F_u^{-1}(m\tau/K(x)) - F_u^{-1}(\tau/K(x))}{\tau(K(x)f_u[F_u^{-1}(\tau/K(x))])^{-1}} &\equiv \int_1^m \frac{f_u[F_u^{-1}(\tau/K(x)])}{f_u[F_u^{-1}(s\tau/K(x)])} ds \\ &\sim \int_1^m s^{-\xi-1} dls = \frac{m^{-\xi}-1}{-\xi}. \end{aligned} \tag{A.3.24}$$

Calculations for the term $\frac{E\eta_t}{\sqrt{\tau T}}$ are analogous, using change of variables:

$$\frac{E\eta_t}{\sqrt{\tau T}} \sim \frac{-1}{2T} E \frac{1}{\mathcal{H}(X_t)} | \frac{m^{-\xi}-1}{-\xi} | z' X_t X_t' z, \text{ uniformly in } t \tag{A.3.25}$$

Combine (A.3.25) and (A.3.19) to conclude. ■

**CLT & LLN under Dependence**

***Lemma A.3.2 (CLT & LLN)***

*Assume the above tail heterogeneity model. Suppose $\{Y_j, X_j\}_{-\infty}^t$ is stationary $\alpha$-mixing triangular sequence.*

*(i) If $\alpha_j = O(j^{-\phi})$, $\phi > 2$, and for any $K$ sufficiently close to 0 or $-\infty$, uniformly for $s, t \geq 1$, and $C > 0$ ($P_t$ denotes conditioning on $\mathcal{F}_t \equiv \sigma(\{Y_j, X_j\}_{-\infty}^t)$):*

$$P_t(U_t \leq K, U_{t+s} \leq K) \leq C P_t(U_t \leq K)^2, \tag{A.3.26}$$

*BC*(3) holds with $\mathcal{G}_{ij} \equiv \mathcal{Q}_X \min(l_i, l_j)/\sqrt{l_i l_j}$. If (A.3.26) is dropped, BC*(3) still holds if $\mathcal{G}$ is nondegenerate, with $\mathcal{G}_{ij} \equiv \lim_T Ep(l_i)p(l_j)$, $p(l) \equiv s_1 - E(s_1|\mathcal{F}_0)$, $s_t \equiv 1[Y_t \leq X_t'\beta(l\tau)] - \tau)X_t$.*

*(ii) If $\alpha_j = O(j^{-\phi}), \phi > \frac{1}{1-\gamma}$, $0 < \gamma < 1$, and $\tau^{1-\frac{2}{\gamma}}/T \to 0$ ($\tau T \to \infty$ not too fast), then BC*(2) holds.*

***Corollary 13.1***

*Combine Theorem 13 with the previous two lemmas to have theorems 4 and 5 in the main text.*

**Remark A.3.2** (Condition (A.3.26) means that extremal events should not cluster too much. It may be refined slightly along the lines of the no-clustering condition of Watts etal.[87](no-covariate case). (A.3.26) is analogous to local no-clustering conditions in [73] (A4.4., p.191) in the context of kernel density estimation.

**Proof of Lemma A.3.2:** (i) $\{W_T(l_i), i \leq m\}$ is of the form suited for the CLT of Robinson[73]. His condition A7.1 (with $q = 0$), A7.2, and A7.3. are satisfied automatically. The $\alpha$-mixing implies $\sum_{j=1}^{\infty} j\alpha_j < \infty$, which implies his condition A3.3. Condition (A.3.26) ensures A7.4.

If (A.3.26) does not hold, we apply theorem of M.I. Gordin (see [38], p.137). It is easy to check, using the assumed mixing condition and the classical Ibragimov inequalities [24], 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) Drop index $l$ from notation, since calculation are identical for all $l > 0$. $Var(G_T(z))$ $= \tau^{-1}O(Var(\mu_1) + 2\sum_{k=1}^{T} E\mu_1\mu_{1+k})$, for $\mu_t$ defined earlier. By binomiality of $\mu_t$ and the calculations analogous to those in Lemma A.3.1 (denote by $\|\cdot\|_{r,P}$ the $L_r(P)$ norm):

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

(for $1/p + 1/r = \gamma \in (0,1)$, $p \geq 1$), using mixing inequality of Ibragimov (see [24]). Conclude $Var(G_T(z)) = o(1)$. ■

## A.4 Model of Congenial Tail Heterogeneity: Properties

The assumptions of this model were stated in part A.2. This section derives the essential properties needed for Theorems 3 and 5 in the main text.

Denote by $M$ any compact subset of $(0,\infty)$ that excludes 1. Let $\mathcal{T}$ be the set of quantile indices $\{\tau : \tau = s\tau', s \in \mathcal{L}\}$, where $\mathcal{L}$ is any compact subset of $(0,\infty)$, and $\tau'$ is the reference index $\searrow 0$. We will make statements that hold uniformly in the given sets. Equivalent (but more ambigious) phrasing is to say "locally uniformly" in given indices (approporiate sets are then deduced from the context). Typically, extreme value literature (e.g.[31]) uses the latter term.

***Lemma A.4.1 ( Properties of Congenial Tail Heterogeneity)***
*Suppose assumption A.2.14 and the linearity [$F_{Y_t}(\tau|X_t = x) = x'\beta(\tau)$ w.p.1].*

*(i)* $K(x)$ *is a function of linear index* $K(x'\mathbf{c})$ *as given before Theorem 12.*

*(ii) Centering constants* $\mathbf{c}(k)$ *are those stated after Theorem 12.*

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

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

$$\beta_1(\tau) - \beta_1(q) \sim F_u^{-1}(\tau), \tag{A.4.27}$$

$$\beta_{-1}(\tau) - \beta_{-1}(q) \sim \mu[F_u^{-1}(m\tau) - F_u^{-1}(\tau)], \tag{A.4.28}$$

*Also if* $\xi \neq 0$, $\beta_{-1}(\tau) - \beta_{-1}(q) \sim \mathbf{c}_{-1} F_u^{-1}(\tau)$.

*(iv)*

$$\frac{(x - \mu_X)'(\beta(\tau) - \beta(q))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \to \begin{cases} (x - \mu_X)' \frac{\mathbf{c}}{m^{-\xi}-1} & \text{if } \xi \neq 0, \\ (x - \mu_X)' \frac{\mathbf{c}}{\ln m} & \text{if } \xi = 0. \end{cases} \tag{A.4.29}$$

*(v)*

$$\frac{x'(\beta(m\tau) - \beta(\tau))}{\mu_X'(\beta(m\tau) - \beta(\tau))} \to \begin{cases} x'\mathbf{c} & \text{if } \xi \neq 0, \\ 1 & \text{if } \xi = 0. \end{cases} \tag{A.4.30}$$

*(vi)*

$$\frac{x'(\beta(m\tau) - \beta(\tau))}{x'(\beta(l\tau) - \beta(\tau))} \to \frac{m^{-\xi} - 1}{l^{-\xi} - 1}. \tag{A.4.31}$$

**Remark A.4.1** The implications of property (iii) is that if $\xi \neq 0$ (type 2 and 3 tails), regression quantile coefficients are approximately like those in location-scale models, far enough in the tails. If $\xi = 0$, the matters are more involved. Take the case of Weibull tail-equivalence, for example: $F_u(z) \sim e^{-(-z)^\nu}$ as $z \searrow -\infty$. Then

$$\begin{aligned} \beta_{-1}(\tau) - \beta_{-1}(q) &\sim \mathbf{c}_{-1}\{(-\ln m - \ln \tau)^{\frac{1}{\nu}} - (-\ln \tau)^{\frac{1}{\nu}}\} \\ &\sim \mathbf{c}_{-1} \frac{1}{\nu} (-\ln \tau)^{\frac{1}{\nu} - 1} \ln m \\ &\sim \mathbf{c}_{-1} \frac{1}{\nu} \ln m \frac{F_u^{-1}(\tau)}{-\ln \tau}. \end{aligned}$$

That is, the *slope* coefficients "dampen" relative to those in the general location-scale model by a logarithmic factor. Hence congeniality in the exponentially tailed cases requires a delicate interference of covariates in the tails.

We need the following facts ( *Caution*: these are stated for the lower quantiles, not high quantiles as more commonly used in the extreme value theory). In the sequel write $F_u \in D(H_\xi)$, if $F_u$ is a cdf in the domain of minimum attraction with the extremal tail index $\xi$. Write $F_u \in \mathcal{R}_\gamma(0)$, if $F_u$ is regularly varying with exponent $\gamma$ at 0. [Use $F_u(z|x)$ to denote $F_{u_t}(z|X_t = x)$, in the sequel.]

**Remark A.4.2 (Facts)** Uniformly in $(m, l, \tau) \in M \times M \times \mathcal{T}$, as $\tau' \searrow 0$,

(i) If $F_1 \sim F_2$ 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 $-\infty$. Suppose function $K(x)$ 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) also 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 $\ell$ is the auxiliary function defined in the main text.

Except (ii), (i)-(iv) can be found in the texts on extreme values – [72], [56], [31]. (ii) holds from (i) pointwise, and uniformly – by the assumed linearity of $F_{u_t}^{-1}(\tau|x)$ in $x$ and compactness of $\mathbf{X}$.

**Proof of Lemma A.4.1** In this proof, term "locally uniformly" is equivalent to "uniformly in $(l, m, \tau, x) \in M \times M \times \mathcal{T} \times \mathbf{X}$, as $\tau' \searrow 0$".

At first, note that $\mu_X'(\beta(\tau) - \beta(q)) \equiv \beta_1(\tau) - \beta_1(q) = F_u^{-1}(\tau|\mu_X) \sim F_u^{-1}(\tau)$, locally uniformly.

$$\begin{aligned}
&\frac{(x-\mu_X)'(\beta(\tau)-\beta(q))}{\mu_X'(\beta(m\tau)-\beta(\tau))} \equiv (x-\mu_X)'A(\tau) \equiv \frac{F_u^{-1}(\tau|x) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \\
&\sim \frac{F_u^{-1}(\tau/K(x)) - F_u^{-1}(\tau)}{F_u^{-1}(m\tau) - F_u^{-1}(\tau)} \text{ locally uniformly} \\
&\sim \frac{(1/K(x))^{-\xi} - 1}{m^{-\xi} - 1} \text{ locally uniformly} \\
&\equiv B(x),
\end{aligned} \tag{A.4.32}$$

which follows from the facts (ii) and (iii). Since $0 < B(x) < \infty$, (A.4.32) implies

$$\left|(x-\mu_X)'A(\tau) - B(x)\right| \to 0 \text{ locally uniformly,} \tag{A.4.33}$$

and that, since $(x-\mu_X)_{-1}$ ranges over a non-degenerate subset of $\mathbb{R}^{d-1}$,[5]

$$A_{-1}(\tau) \to \kappa(m) \text{ locally uniformly,} \tag{A.4.34}$$

where $\kappa(m)$ is a constant vector. Hence (A.4.33) and (A.4.34) imply that $B(x)$ is *affine* in $x$ on $\mathbf{X}$.

If $\xi = 0$, conclude $B(x) = -\ln K(x)/\ln m = \mathbf{c}(x-\mu_X)/\ln m$. Hence $K(x) = e^{-x'\mathbf{c}}$ on $\mathbf{X}$ s.t. $\mu_X'\mathbf{c} = 0$, i.e. $\mathbf{c}_1 = 0$. If $\xi \neq 0$, conclude $B(x) = (K(x)^\xi - 1)/(m^{-\xi}-1) = \mathbf{c}(x-\mu_X)/(m^{-\xi}-1)$. Hence $K(x) = (x'\mathbf{c})^{1/\xi}$ on $\mathbf{X}$, s.t. $\mu_X'\mathbf{c} = 1$, i.e. $\mathbf{c}_1 = 1$. 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) of the lemma.

Claim (iv) is verified by substituting the forms of $K(x)$ found into (A.4.32). Claim (iii) follows directly from equations (A.4.32), (A.4.34), 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) is next. If $\xi \neq 0$, by claim (iii), uniformly in $k$ in any compact subset of $(0,\infty)$ as $T \to \infty$

$$a_T(\beta(\tfrac{k}{T}) - \beta(q)) \sim a_T\mathbf{c}F^{-1}(\tfrac{k}{T}) = \mathbf{c}F_u^{-1}(\tfrac{k}{T})/F_u^{-1}(\tfrac{1}{T}) \sim k^{-\xi}\mathbf{c}, \tag{A.4.35}$$

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

$$\begin{aligned} &a_T(\beta(\tfrac{k}{T}) - \beta(q) - b_T\mathbf{e}_1) \sim \\ &\frac{1}{\ell(F^{-1}(1/T))}\left[\mathbf{c}\left(F_u^{-1}(e\tfrac{k}{T}) - F_u^{-1}(\tfrac{k}{T})\right) + \mathbf{e}_1\left(F_u^{-1}(\tfrac{k}{T}) - F_u^{-1}(\tfrac{1}{T})\right)\right] \\ &\to \mathbf{c}\ln e + \mathbf{e}_1 \ln k = \mathbf{c} + \mathbf{e}_1\ln k. \end{aligned} \tag{A.4.36}$$

Claim (v) holds pointwise in $x$ by the facts (ii) and (iii). Since the ratio on the l.h.s. in (A.4.30) is linear in $x$ and $\mathbf{X}$ is compact, it also holds uniformly in $x \in \mathbf{X}$ [$span(\mathbf{X}) = \mathbb{R}^d$ by nondegeneracy of $\mathbf{X}$; thus any $x$ in $\mathbf{X}$ is a linear combination of some points $x_1, ..., x_d$ in $\mathbf{X}$. So it suffices that (v) holds for points $x_1, ..., x_d$ only, by compactness of $\mathbf{X}$.]

Combine fact (iii) with claim (v) to have claim (vi). ■

[5] $(x-\mu_X)_1 = 0$ by assumption.

## A.5 Design Conditions for Extreme Ranks

This design condition is a general setting that insures essential uniqueness in the limit program in the extreme rank asymptotics. It has some other interesting implications about the break-points.

**Condition BC($3^\bullet$) ( That Guarantees Essential Uniqueness)** (a) For any set $\mathbb{X}$ in $\mathbf{X}$ s.t. $\int_{\mathbb{X}} dF_{\mathcal{X}} > 0$ and any $\kappa_1, \epsilon > 0$, there is $\kappa_2$ large s.t.

$$\mathbf{N}([-\infty, \kappa_2] \times \mathbb{X} \cap E) > \kappa_1 \text{ w. pr. } \geq 1 - \epsilon\ . \tag{A.5.37}$$

(b) $F_{\mathcal{X}}$ is non-degenerate in $\mathbb{R}^d$, (c) for at least one $i$, $F_{\mathcal{X}_i}$ is absolutely continuous (if $d \geq 2$), and $\{J_i\}$ are continuously distributed cond'l on $\{\mathcal{X}_i\}$.

Assumptions (a)- (c) hold for all specific PRM that we have considered, except that presence of one continuous covariate if $d \geq 2$, in assumption (c), shortens the argument, but may be relaxed.

**Remark A.5.1** Nondegeneracy (b) implies that (see Portnoy and Jureckova[68]) for some positive constants $\delta$ and $K$ $\exists$ sets $R_i$, $i \leq K$ (that cover $\mathbb{R}^d$) s.t. for any $c : \|c\|_2 > 1$, $\exists i$ s.t. $\int_{R_i} dF_{\mathcal{X}} > 0$ and $x'c > \delta \|c\|_2$.

***Lemma A.5.1 (Essential Uniqueness, Tightness, and Break-Points)***
*Under condition BC($3^\bullet$), (i) for any $\mathcal{K} = [K_1, K_2]$, $\sup_{k \in \mathcal{K}} \|Z(k)\| = O_{P^*}(1)$, (ii) $\#\mathcal{K}_b = O_{P^*}(1)$ for $E_1$ and $E_3$. (and for $E_2$ if $d = 1$), (iii) $k \notin \mathcal{K}_b$ w.p.1. if $d \geq 2$, and $Leb(\mathcal{K}_b) = 0$, if $d = 1$.*

**Proof of Lemma A.5.1(i)** Select $z^f \in \mathbb{R}^d$ s.t.

$$\sup_{k \in \mathcal{K}} \left| Q_\infty(z^f, k) \right| = O_P(1). \tag{A.5.38}$$

(A.5.38) is possible as shown in the Proof of Theorem 10 and because $k\mu_X$ enters $Q_\infty$ linearly. By such linearity property and convexity, if $z_1$ and $z_2$ are s.t. (A.5.38) holds, (A.5.38) also holds for any $z_3$ in 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, and $\delta(k) > M$ for any $k$. $k$ can be aribitrary and it is not fixed. Only the cases when $Q_\infty(z(k), k)$ is finite a.s. are interesting (otherwise we have what we need immediately). By convexity for all $k$

$$\frac{M}{\delta(k)} \big( Q_\infty(z(k), k) - Q_\infty(z^f, k) \big) \geq Q_\infty(z^*(k), k) - Q_\infty(z^f, k), \tag{A.5.39}$$

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

$$P_*\big(\inf_{k\in\mathcal{K}} Q_\infty(z^*(k),k) > K\big) \geq 1-\epsilon. \tag{A.5.40}$$

(A.5.40), combined with (A.5.38), implies r.h.s. of $(A.5.39) > \delta > 0$ w.p. arbitrarily close to 1 for $M$ large enough. This verifies claim (a) of the lemma. It remains to show (A.5.40). For given direction $v(k)$, $\mu_X' z^*(k)$ is either (i) increasing to $\infty$ at rate $M$ or constant ($z_1^f(k)+M$ or $z_1^f(k)$ ) (ii) decreasing to $-\infty$ at rate $M$, as $M\to\infty$. Case (ii) obviously implies (A.5.40). Hence we look only at (i). Fix some $\kappa_1 > 1$. We shall prove that Condition BC($3^\bullet$) implies that for any $K$ and $\epsilon > 0$, uniformly in $k\in\mathcal{K}$

$$-\kappa_1 M + \int_E l(u,x'z^*(k))d\mathbf{N} > K \text{ w. pr. } \geq 1-\epsilon\ ,\ \text{as } M\to\infty, \tag{A.5.41}$$

and therefore (A.5.40). Thus it suffices to show that uniformly in $k$ for some $\kappa_2 > \kappa_1$

$$\int_E l(u,x'z^*(k))d\mathbf{N} > \kappa_2 M + \kappa_3 \text{ w. pr. } \geq 1-\epsilon\ ,\ \text{as } M\to\infty. \tag{A.5.42}$$

By Remark A.5.1 we have that for any direction $v(k), k\in\mathcal{K}$, there is $\mathbb{X}_{i(k)} = \{x\in\mathbf{X} : x'z^*(k) \geq \kappa_4 M\}$ s.t. $\int_{\mathbb{X}} dF_{\mathcal{X}} > 0$ and only $K$ such sets $\mathbb{X}_i$ correspond to all possible directions $v(k)$ and $k$. Hence for any $(z^*(k),k)$ s.t. as $M\to\infty$ (for some $i$)

$$\begin{aligned}\int_E l(u,x'z^*(k))d\mathbf{N} &\geq \int_{E\cap[-\infty,\kappa_5]\times\mathbb{X}_i} (u-x'z^*(k))^- d\mathbf{N} \\ &\geq \mathbf{N}(E\cap[-\infty,\kappa_5]\times\mathbb{X}_{i(k)})(\kappa_4 M-\kappa_5).\end{aligned} \tag{A.5.43}$$

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

**Proof of Lemma A.5.1(ii)-(iii)** Let $\mathcal{B}(M)$ be event s.t. $\mathbf{N}(E\cap[-\infty,\kappa_5]\times\mathbb{X}_i) > \kappa_2/\kappa_4$ for all $i$. This is a measurable event that implies $\sup_{k\in\mathcal{K}}\|Z(k)\|\in B(M)$. Conditionally on $\mathcal{B}(M)$, one has:

$$Z(k) = \arg\min_{z\in B(M)} Q_\infty(z,k) = \arg\min_{z\in\mathbb{R}^d} -k\mu_X' z + \sum_{t:J_t\leq\kappa}(J_t-\mathcal{X}_t'z)^-. \tag{A.5.44}$$

where $\kappa = \max_{z\in B(M),x\in\mathbf{X}} x'z < \infty$. Hence $\mathbf{N}([-\infty,\kappa]\times\mathbf{X}) \equiv N_k$ is finite a.s. for $E_1$ and $E_3$. Therefore the r.h.s. of (A.5.44) is just a 'finite-sample' linear program that under BC($3^\bullet$) (c) has break-point number $\#\mathcal{K}_b \leq N_k!/(d!(N_k-d)!) \leq N_k^d$ (see e.g. [55], [67]). Hence for any $\epsilon > 0$, there is $K$ and $M$ large enough s.t.

$$P^*\big(\#\mathcal{K}_b > K\big) \leq P^*\big(\#\mathcal{K}_b > K|\mathcal{B}(M)\big) + P(\mathcal{B}(M)^c) \leq P\big(N_k^d > K|\mathcal{B}(M)\big)+\epsilon \leq 2\epsilon.$$

This shows (ii). To show (iii), either use (ii) or note that distributional assumptions (c) and the gradient conditions (see main text) for the linear programming problem (A.5.44) insure that $k \in \mathcal{K}_b$ w.p. 0, if $d \geq 2$.[6] [ In case of $E_2$, BC(3)(2) insures that $Q_\infty(Z(k), k) = O_p(1)$ for all $k$, and so the gradient conditions apply too)] In case of $d = 1$, examining the gradient condition, observe that $\mathcal{K}_b$ consists of integers $\{1, 2, ...\} \cap \mathcal{K}$. ■

[6] See e.g. [67]

# Appendix B

# Appendix for Chapter 4

## B.1 Extreme Ranks

### B.1.1 Useful Lemmas

Let $\mathbf{H}$ be a real Hilbert space with inner product $\langle .,. \rangle$ and complete orthonormal system $\{e_j, j \in J\}$, so that for any $x \in \mathbf{H}$, $x = \sum_j \langle X_j, e_j \rangle e_j$ . Let $(\Omega, \mathcal{F}, P)_n$ be an indexed by $n$ probability space. Net $X_n : \Omega \mapsto \mathbf{H}$ is *asymptotically finite dimensional (a.f.d.)* if for all $\delta > 0$ and $\epsilon > 0$, there exists finite collection $I \subset J$ s.t.

$$\overline{\lim}_n P^* \big( \sum_{j \notin I} \langle X_n, e_j \rangle^2 > \delta \big) < \epsilon.$$

***Lemma B.1.1***
*$X_n$ is asymptotically tight (a.t.) if $X_n$ is* a.f.d. *and $\langle X_n, e_j \rangle$ is* a.t. *for all $j \in J$.*

***Lemma B.1.2***
*Let $X_n$ be a.t., then it is asymptotically measurable iff $\langle X_n, e_j \rangle$ is asymptotically measurable for all $j \in J$.*

***Lemma B.1.3***
*$X_n \Rightarrow X$ in $\mathbf{H}$, where $X$ is a tight Borel measurable map, iff $X_n$ is* a.f.d. *and net $\langle X_n, h \rangle \xrightarrow{d} \langle X, h \rangle$ for all $h \in \mathbf{H}$.*

Lemmas B.1.1-B.1.3 are those of [85], p. 49-50.

### B.1.2 Asymptotics for Extreme Ranks

The basic settings are as those in the appendix for chapters 2 and 3.

***Assumption 7***
*The following hold:*

- *Basic Conditions 1,2,3$^\bullet$, 4 (stated in appendix A)*
- *For space $E_2$:*

$$\sup_{x\in\mathbf{X},k\in\mathcal{K}} x'Z_T(k) < 0 \text{ w.p. } \to 1, \text{ for } T \leq \infty. \tag{B.1.1}$$

**Remark B.1.1** Assumptions BC(1),(2),(4) have been justified in the previous chapters. Assumption BC(3$^\bullet$) and condition (B.1.1) replace BC(3) employed before. BC(3$^\bullet$) has been justified as a general assumption – it holds for the model of the congenial tail heterogeneity. Condition (B.1.1) holds trivially when $d = 1$, and otherwise when covariates are discrete (which for all practical purposes may suffice). Extending it to general situations seems to be a technical matter.

Since processes are vector- valued, $L_2(K, \mu_L)$ is constructed in a standard manner as a product of the usual $L_2$ spaces that accommodate real-valued functions. Hence, for any vector valued $f, h$ in $L_2(K, \mu_L)$, inner product $\langle f, h\rangle$ is vector valued with components $\langle f_i, h_i\rangle = \int_{\mathcal{K}} f_i(k)h_i(k)dk$. A norm $\|\cdot\|_{2,\mu_L}$ is defined, for example, as the maximum of the norms of the components: $\|f\|_{2,\mu_L} = \max_i\langle f_i, f_i\rangle^{1/2}$.

***Theorem 14 (Extremal Process under Extreme Ranks)***
*Given assumption 7, $Z_T(\cdot) \Rightarrow Z_\infty(\cdot)$ in $L_2(\mathcal{K}, \mu_L)$.*

***Corollary 14.1***
*Theorem 6 holds as a special case of Theorem 14.*

**Proof of Theorem** 14. An important ingredient of the proof is the following lemma (proof follows after).

***Lemma B.1.4***
*Under assumption 7*

*(i)* $\sup_{x\in\mathbf{X},k\in\mathcal{K}} \|Z_T(k)\|_\infty = O_{P^*}(1)$,

*(ii)* $\sup_{x\in\mathbf{X},k\in\mathcal{K}} \|Z_\infty(k)\|_\infty = O_{P^*}(1)$,

*(iii)* $\#\mathcal{K}_{bT} = O_{P^*}(1)$,

*(iv)* $\#\mathcal{K}_b = O_{P^*}(1)$,

*where $\#\mathcal{K}_{bT}$ and $\#\mathcal{K}_b$ are numbers of break-points in $\mathcal{K}$ of the finite-sample and infinite sample optimization programs (see appendix A).*

Space $L^2(\mathcal{K}, \mu_L)$ is a Hilbert space functions with a complete orthonormal system $\{e_j, j \in J\}$ [ defined for example by $\sin(a_j \cdot)$ and $\cos(b_{j'} \cdot)$ functions with properly chosen coefficients $a_j$ and $b_{j'}$, see any text on Hilbert spaces.] with the usual inner product and norm. It is helpful to note that for any $f = (f_1, ..., f_d)$ , $\|f\|^2_{2,\mu_L} = \max_i \int_{\mathcal{K}} |f_i(k)|^2 dk$ $= \max_i \sum_{j \in J} \langle f_i, e_{ji} \rangle^2$.

By Lemma B.1.4 $Z_T$, $Z_\infty$ belong to $L^2(\mathcal{K}, \mu_L)$ *w.p.* $\to 1$. Proof of the theorem now proceeds in three steps:

- $Z_T(\cdot)$ is a.f.d.,
- $Z_\infty(\cdot)$ is tight and Borel measurable random map $\Omega \mapsto L^2(\mathcal{K}, \mu_L)$,
- $\langle Z_T, h \rangle \xrightarrow{d} \langle Z_\infty, h \rangle$ for every $h \in L^2(\mathcal{K}, \mu_L)$.

These imply the theorem by Lemma B.1.3. It only remains to show the steps.

1. Working directly with $Z_T(\cdot)$ appears to be cumbersome, since $Z_T$ is a piece-wise constant function with randomly spaced break-points. Number of break-points, $\#\mathcal{K}_{bT}$ is also random albeit stochastically bounded. Yet $Z_T(\cdot)$ is approximable by a well-behaved piece-wise constant approximation map with a finite number of regularly spaced break-points. Construct the approximation operator as $A^\delta$ mapping a given $f$ to

$$(A^\delta f)(\cdot) = \sum_{i \in I(\delta)} f(k_i) 1\{\cdot \in [k_i, k_{i+1})\},$$

where $k_{i+1} = k_i + \delta$, $k_1 = \underline{k}$, $I(\delta) = \{1, 2, ..., \mathcal{I}(\delta)\}, \mathcal{I}(\delta) \leq (\bar{k} - \underline{k})/\delta + 1$.

Then $\overline{\lim}_T P^* \Big[ \|A^\delta Z_T - Z_T\|_{2,\mu_L} > \epsilon \Big]$

$$\begin{aligned} &\leq \overline{\lim}_T P^* \Big[ \Big( \sum_{i \in I(\delta)} 1\{k_b \in [k_i, k_{i+1}]\} \Big) \times \sup_{k \in \mathcal{K}} \|Z_T(k)\|_2 \times \delta > \epsilon \Big] \\ &\leq \overline{\lim}_T P^* \Big[ \#\mathcal{K}_{bT} \times \sup_{k \in \mathcal{K}} \|Z_T(k)\|_2 \times \delta > \epsilon \Big] \leq \eta. \end{aligned} \tag{B.1.2}$$

for any given $\eta > 0$, by choosing $\delta$ small, by Lemma B.1.4. [Here $1\{k_b \in [k_i, k_{i+1}]\}$ indicates whether a break-point is present in the interval $[k_i, k_{i+1}]$.

The approximating map $A^\delta Z_T(\cdot)$ is clearly a.f.d. for any $\delta > 0$, since it is a piece-wise constant map defined by a finite number of stochastically bounded variables $\{Z_T(k_i), i \in I(\delta)\}$. Hence $Z_T(\cdot)$ is itself a.f.d. in view of (B.1.2).

2. $Z_\infty(\cdot)$ is clearly finite dimensional by lemma B.1.4 and the argument of step 1.

Write $\langle Z_\infty, e_j\rangle = \langle A^\delta Z_\infty, e_j\rangle + \langle A^\delta Z_\infty - Z_\infty, e_j\rangle$. By Bessel inequality for all $j$ in $J$, $P^*\Big[\|\langle A^\delta Z_\infty - Z_\infty, e_j\rangle\|_{2,\mu_L}\Big]$

$$\leq P^*\Big[\|A^\delta Z_\infty - Z_\infty\|_{2,\mu_L}\|e_j\|_{2,\mu_L} > \epsilon\Big] \leq \eta \tag{B.1.3}$$

for any given $\eta > 0$ by choosing $\delta$ small, by Lemma B.1.4 and the argument of equation (B.1.2) (note $\|e_j\|_{2,\mu_L} = 1$). Thus $\langle A^\delta Z_\infty - Z_\infty, e_j\rangle$ is $O_P(1)$ [ tight in $\mathbb{R}^d$ ]. $\langle A^\delta Z_\infty, e_j\rangle$ $= O_P(1)$, since $\{Z_\infty(k_i), i \in I(\delta)\}$ is $O_P(1)$ by lemma B.1.4. Hence $Z_\infty$ is tight by lemma B.1.1.

Note that $\langle A^\delta Z_\infty, e_j\rangle$ is measurable and term $\langle A^\delta Z_\infty - Z_\infty, e_j\rangle$ can be made arbitrarily small in outer probability by choosing small $\delta$, by (B.1.3). Hence $\langle Z_\infty, e_j\rangle$ is measurable and thus $Z_\infty$ is measurable by lemma B.1.2.

3. For any $h \in L_2(\mathcal{K}, \mu_L)$, $\langle Z_\infty, h\rangle = \langle A^\delta Z_\infty, h\rangle + \langle A^\delta Z_\infty - Z_\infty, h\rangle$, $\langle Z_T, h\rangle = \langle A^\delta Z_T, h\rangle +$ $\langle A^\delta Z_T - Z_T, h\rangle$. $\langle A^\delta Z_T, h\rangle = \sum_{i\in I(\delta)} Z_T(k_i) \int_{k_i}^{k_{i+1}} h(l)dl \xrightarrow{d} \sum_{i\in I(\delta)} Z_\infty(k_i) \int_{k_i}^{k_{i+1}} h(l)dl$ $= \langle A^\delta Z_\infty, h\rangle$, by appendix A. Remainder terms $\langle A^\delta Z_\infty - Z_\infty, h\rangle$ and $\langle A^\delta Z_T - Z_T, h\rangle$ can be made arbitrarily small in outer probability by choosing $\delta$ small, by the argument as in (B.1.2) and (B.1.3). Conclude $\langle Z_T, h\rangle \xrightarrow{d} \langle Z_\infty, h\rangle$ for any $h \in L^2(\mathcal{K}, \mu_L)$. ■

**Proof of lemma B.1.4**. At first, claims (ii) and (iv) are shown in Appendix A (for type space $E_2$, the proof needs in addition the no-crossing condition assumed here). Claims (i) and (ii) will follow similarly from convexity of objective $Q_T(z, k)$ in $z$ and $k$ and hence uniform convergence to $Q_\infty(z, k)$ over suitable compacta.

## B.2 Intermediate Ranks

The definitions of outer/inner measures and equicontinuity may be found in [42], [41], [64], and [86].

### B.2.1 Asymptotics under Intermediate Ranks

Notation that we do not define here is defined in the main text. We like to study the convergence of map:

$$Z_T = \left\{ a_T(l)(\hat{\beta}(l\tau) - \beta(l\tau)), l \in \mathcal{L} \right\}$$

for the normalization factor $a_T(l)$ to be defined, and $\mathcal{L} \equiv [\underline{l}, \bar{l}]$, for $\bar{l} \geq \underline{l} > 0$.

The process $Z_T(\cdot)$ is not measurable w.r.t. to the ball $\sigma$-field of $\ell^\infty(\mathcal{L})$, but the conjectured weak limit $Z_\infty(\cdot)$, defined as the finite-dimensional weak, is a measurable tight Gaussian process. Thus choosing $\ell^\infty$ as a metric space, to embed our process in, is appropriate; and usage of [41] weak convergence theory will be fruitful.

The following are the general assumptions that rather flexibly accommodate heterogeneous situations and dependent data. [Notation that we do not define here is defined in previous chapters.]

***Assumption 8***

- ***BC*(1), BC*(2), BC*(3)*** *stated in appendix A*
- ***BC*(4)***: *(a) map $W_T(\cdot)$ is s.e., (b) map $\{G_T(l,z), l \in \mathcal{L}, z \in K\}$ is s.e. for any compact non-empty set $K \subset \mathbb{R}^d$.*

**Remark B.2.1** 1. Assumptions BC*(1), BC(2), BC*(3) are shown to be general and plausible in the previous chapters (appendix A). These require applicability of certain analytical properties, CLTs and LLNs. Appendix A has shown how these assumptions hold for basic model of tail-homogeneity as well as the model of the congenial tail heterogeneity (under i.i.d. as well as reasonable serial dependence conditions). 2. BC*(4) is the added assumption requiring the asymptotic tightness of the processes that forms the Jacobian and "Hessian" of the objective function. This, combined with BC*(1)-BC*(3), requires the uniform LLN and CLT to apply. We shall demonstrate that BC*(4) is plausible for the basic model of tail homogeneity as well as for the model of the congenial tail heterogeneity.

***Theorem 15 (Convergence in $\ell^\infty(\mathcal{L})$)***
*Under conditions BC*(1)- BC*(4)*

$$Z_T(\cdot) \Rightarrow Z_\infty(\cdot) \quad \text{in } \ell^\infty(\mathcal{L}), \text{ as } \tau \to 0, \tau T \to \infty$$

*where $Z_\infty(\cdot)$ is a centered Gaussian process, with covariance*

$$E(Z_\infty(l_1)Z_\infty(l_2)') = J^{-1}(l_1)\mathcal{G}(l_1,l_2)J^{-1}(l_2)', \tag{B.2.4}$$

*where $J(l)$ is defined in appendix A, and $\mathcal{G}(l_1,l_2)$ is the asymptotic covariance between $W_T(l_1)$ and $W_T(l_2)$.*

***Corollary 15.1***
*Theorem 7 follows in view of sufficient conditions provided in appendix for Chapter 3 (BC*(1) – BC*(3)) and the section below (BC*(4)).*

**Proof of Theorem 15** The objective function that $\hat{Z}_T(l) \equiv a_T(l)\Big(\hat{\beta}(l\tau)-\beta(l\tau)\Big)$ minimizes can be written as:

$$\widetilde{Q_T}(Z,\tau l) \equiv \sum_t \rho_{(\tau l)}\left(Y_t - X_t'\beta(\tau l) - \frac{X_t'Z}{a_T(l)}\right) - \sum_t \rho_{(\tau l)}\left(Y_t - X_t'\beta(\tau l)\right)$$

The subtraction of the latter term does not affect optimization. Also define the normalized objective as:

$$Q_T\left(l,z\right) \equiv \frac{1}{\sqrt{(\tau l)T}} a_T(l) \times \widetilde{Q_T}(Z,\tau l).$$

Rewrite:

$$\begin{aligned} Q_T\left(l,z\right) \equiv & \frac{-1}{\sqrt{(\tau l)T}}\Big[\sum_t \left(\tau l - 1\{Y_t \le X_t'\beta(\tau l)\}\right) X_t'\Big] z \\ & - \frac{1}{\sqrt{(\tau l)T}}\sum_t \mu_t(l,z)X_t'z + \frac{1}{\sqrt{(\tau l)T}}\sum_t \mu_t(l,z)[Y_t - X_t\beta(\tau l)]a_T(l) \\ \equiv & \; W_T(l)'z \; + \; G_{1T}(l,z) \; + \; G_{2T}(l,z). \end{aligned}$$

where $\mu_t(l,z) \equiv [1\{Y_t \le X_t'\beta(\tau l)\} - 1\{Y_t \le X_t'(\beta(\tau l) + z/a_T(l))\}]$.

By BC*(1) and (4), for any compact set $K$ in $\mathbb{R}^d$:

$$\sup_{(l,z)\in\mathcal{L}\times K}\left\|G_{1T}(l,z) + G_{2T}(l,z) - z'\frac{1}{2}J(l)z\right\| \xrightarrow{p} 0. \tag{B.2.5}$$

(B.2.5) holds pointwise by BC*(2); BC*(4) strengthens it to the uniform convergence in probability [85]. It also implies that $J(l)$ must be continuous on $\mathcal{L}$ by Theorem 1.5.7 in [85]. Furthermore, by BC*(1) and BC*(4)(a), Theorems 1.5.4 and 1.5.7 of [85], in $\ell^\infty(\mathcal{L})$

$$W_T(\cdot) \Rightarrow W_\infty(\cdot), \text{ hence } \sup_{l\in\mathcal{L}}\|W_T(l)\|_2 = O_{P*}(1), \tag{B.2.6}$$

and $W_\infty(\cdot)$ is a tight Gaussian process. Therefore, $Q_T(l,z) = W_T(l)'z + z'\frac{1}{2}J(l)z + \bar{r}_T(l,u)$, where $\sup_{(l,z)\in\mathcal{L}\times K}\|\bar{r}_T(l,u)\|_\infty = O_{P*}(1)$. To simplify notation in the sequel, restate this as follows: $\hat{u}_T(l) \equiv J^{1/2}(l)Z_T(l)$ minimizes objective $V_T(l,u) \equiv Q_T(l, J^{-1/2}(l)u)$, hence for $\tilde{u}_T(l) \equiv J^{-1/2}(l)W_T(l)$

$$\begin{aligned} V_T(l,u) &\equiv \frac{1}{2}\|u - \tilde{u}_T(l)\|_2^2 - \frac{1}{2}\|\tilde{u}_T(l)\|_2^2 + r_T(l,z), \\ \Delta_T &\equiv \sup_{(l,u)\in\mathcal{L}\times U}\|r_T(l,z)\|_2^2 \xrightarrow{p} 0, \end{aligned} \tag{B.2.7}$$

(where $U$ is any compact subset of $\mathbb{R}^d$). Our goal is to show

$$\sup_{l\in\mathcal{L}}\|\hat{u}_T(l) - \tilde{u}_T(l)\|_2 \xrightarrow{p} 0. \tag{B.2.8}$$

If $\mathcal{L}$ is finite, (B.2.8) holds by the standard convexity arguments (see e.g. [65]). In our case $\mathcal{L}$ is an interval $[\underline{l}, \bar{l}]$, and we need to make an extension of such convexity arguments to claim the uniformity. This extension is simple.

Take a ball $B_\delta(l)$ of radius $\delta$ centered at $\tilde{u}_T(l)$ for every $l$. To prove (B.2.8), it suffices to show that $\hat{u}_T(l)$ can not leave $B_\delta(l)$ for any $l \in \mathcal{L}$ w.p. $\to 1$, for any fixed $\delta > 0$. Hence consider behavior of $V_T(l, u(l))$ when $\hat{u}_T(l)$ is outside $B_\delta(l)$ for any arbitrary $l$ ($l$ is not fixed here) Let $u(l) = \tilde{u}_T(l) + c(l)v(l)$, where $c(l) \geq \delta$ for all $l > 0$ and $v(l)$ is a unit vector (of norm 1). Let $u_T^*(l)$ denote he boundary point of $B_\delta(l)$ that is on the line segment connecting $\tilde{u}_T(l)$ and $u(l)$. By (B.2.6) $U$ can be chosen to contain $B_\delta(l)$ for all $l$ with probability arbitrarily close to 1. Therefore, using convexity,

$$\begin{aligned} (\delta/c(l))V_T(l,u(l)) + (1-\delta/c(l))V_T(l,\tilde{u}_T(l)) &\geq V_T(l, u_T^*(l)) \\ &\geq \delta^2/2 - \|\tilde{u}_T(l)\|_2^2 - \Delta_T \\ &\geq \delta^2/2 + V_T(l,\tilde{u}_T(l)) - 2\Delta_T. \end{aligned}$$

But $\delta^2/2 - 2\Delta_T > \epsilon$ w.p. $\to 1$ for some $\epsilon > 0$ by (B.2.7). Hence w.p. $\to 1$ uniformly in $l \in \mathcal{L}$, and any $\delta > 0$:

$$V_T(l,u(l)) - V_T(l,\tilde{u}_T(l)) > c(l)\epsilon/\delta \geq \epsilon,$$

Therefore, w.p.$\to 1$, for no $l$ can the optimum leave $B_\delta(l)$. That is, for any $\delta > 0$

$$P^*\left[\sup_{l\in\mathcal{L}}\left\|\tilde{u}_T(l) - \hat{u}_T(l)\right\|_2 \geq \delta\right] \to 0. \tag{B.2.9}$$

(B.2.9) implies $\sup_{l\in\mathcal{L}}\left\|Z_T(l) - J^{-1}(l)W_T(l)\right\|_2 \xrightarrow{p} 0$. The theorem now follows. ■

### B.2.2 Sufficient Conditions for Stochastic Equicontinuity

This section verifies that BC*(4) holds for the case of *i.i.d.* data in the basic model of tail homogeneity as well as for the congenial tail heterogeneity. Extensions to dependent data can be carried out along the same lines using the Bernstein-Hoeffding type or moment maximal inequalities (see [29] for many of them).

***Lemma B.2.1 (Sufficient Condition for BC*(4))***
*(i) Suppose $\{Y_t, X_t\}$ is an* i.i.d. *triangular array, then BC*(4)(a) holds. (ii) Suppose further that model of the congenial tail heterogeneity holds, then BC(4)(b) holds.*

Before proceeding, recall the definitions of the bracketing numbers and entropy – cf. [85]. Fix a collection $\mathcal{F}$ of functions $f$. $L_2(P)$ norm of $\mathbb{R}^d$-valued $f = (f_1, ..., f_d)$ is defined as $\|f\|_{2,P} \equiv \max_{i\leq d}[E|f_i|^2]^{1/2}$. Given two functions $\underline{b}, \bar{b}$, the bracket $[\underline{b}, \bar{b}]$ is the set of all functions $f$ in $\mathcal{F}$ with $\underline{b} \leq f \leq \bar{b}$. An $\epsilon$ bracket in $L_2(P)$ is $[\underline{b}, \bar{b}]$ with $\|\underline{b} - \bar{b}\|_{2,P} \leq \epsilon$. The $L_2(P)$ bracketing number $N_{[]}(\epsilon, \mathcal{F})$ is the minimum number of brackets needed to cover $\mathcal{F}$. The bracketing entropy $J_{[]}(\epsilon, \mathcal{F})$ equals $\int_0^\delta \ln^{1/2} N_{[]}(\epsilon, \mathcal{F})d\epsilon$.

**Proof of Lemma B.2.1 (i) :**

We show here that $\{W_T(l), l \in \mathcal{L}\}$ is s.e. w.r.t semimetric $\rho(l_1, l_2) = \|\eta(l_1) - \eta(l_2)\|_{2,P}$, where $\eta(l) \equiv X_t 1(Y_t \leq X_t'\beta(\tau l))/\sqrt{\tau}$ [ and $(\mathcal{L}, \rho)$ is totally bounded.] First, we see that:

(i) there is an envelope function $F_T$ s.t. $|\eta(l)| \leq F_T$ for all $l$ in $\mathcal{L}$:

- $EF_T^2 = O(1)$
- $EF_T^2 1\{F_T > \epsilon\sqrt{T}\} \to 0, \forall \epsilon > 0.$

(ii) $J_{[]}(\epsilon, \mathcal{F}_T) \to 0$ as $\delta \to 0$, where $\mathcal{F}_T = \{\eta(l), l \in \mathcal{L}\}$.

To show (i), choose the envelope function $F_T = 1(y \leq x'\beta(\tau\bar{l}))m\mathbf{1}$, where $m = \sup_{x\in\mathbf{X}} \|x\|_\infty$ and $\mathbf{1}$ is a $d$-vector of ones. By monotonicity of quantile function $x'\beta(\cdot)$ and of indicator $1(\cdot)$, $F_T$ envelopes every $f$ in $\mathcal{F}_T$. It is also straightforward to see that $EF_T^2 = \bar{l}m\mathbf{1}$ and $EF_T^2 1\{F_T > \epsilon\sqrt{T}\} = 0$ once $\tau T$ is large enough. To show (ii), it is straightforward that by monotonicity and definition of $1(y \leq x'\beta(\cdot))x/\sqrt{\tau}$ we only need $O(1/\delta^2)$ brackets to cover $\mathcal{F}_T$. [ Thusly $(\mathcal{L}, \rho)$ is totally bounded.] Hence (ii) follows.

Given (i) and (ii), by lemma 19.3.4 in [84] and construction of $\rho$:

$$E \sup_{\rho(l_1,l_2)<\delta} \|W_T(l_1) - W_T(l_2)\|_2 \leq O\Big[J_{[]}(\delta, \mathcal{F}_T) + \frac{EF^2 1\{F_T > K(\delta)\sqrt{T}\}}{K(\delta)}\Big] \tag{B.2.10}$$

where $K(\delta) > 0$ for every $\delta > 0$. The r.h.s. of (B.2.10) can be made arbitrarily small by first letting $T \to \infty$, followed by choosing $\delta > 0$ small, by (i) and (ii). ■

**Proof of Lemma B.2.1 (i) :** $G_T(l,z) = G_{1T}(l,z) + G_{2T}(l,z)$. Since sum of two s.e. processes is s.e, the terms can be considered separately. We consider term $G_{1T}(\cdot,\cdot)$ only, since proof for $G_{2T}(\cdot,\cdot)$ is very similar. Further write

$$\begin{aligned}G_{1T}(l,z) &= \frac{1}{\sqrt{\tau T}}\sum_t 1\Big[X_t'\beta(\tau l) \le Y_t \le X_t'(\beta(\tau l) + \frac{z}{a_T})\Big]X_t'z \\ &- \frac{1}{\sqrt{\tau T}}\sum_t 1\Big[X_t'(\beta(\tau l) + \frac{z}{a_T}) \le Y_t \le X_t'(\beta(\tau l))\Big]X_t'z \\ &\equiv \frac{1}{\sqrt{T}}\sum_t v_t(l,z)X_t'z + \frac{1}{\sqrt{T}}\sum_t w_t(l,z)X_t'z.\end{aligned}$$

We give proof for the first term only, since proof for the second term differs only in notation. Next step is to show that properties (i) and (ii) in the preceding proof hold for the class of functions

$$\mathcal{F}_T = \{v_1(l,z)\cdot X_1 z, (l,z) \in \mathcal{L}\times K\}$$

Define semimetric $\rho((l,z),(l',z')) = \|v_1(l,z)X_1'z - v_1(l',z')X_1'z\|_{2,P}$. Choose envelope function

$$F_T(x,y) = 1(x'\beta(\tau\underline{l}) \le y \le x'\beta(\tau\bar{l}) + c/a_T)c/\sqrt{\tau}$$

with $c = \sup_{x\in\mathbf{X},z\in K}|x'z|$. It envelopes any $f$ in $\mathcal{F}_T$ by monotonicity of quantile function $x'\beta(\cdot)$ and of $1(\cdot)$. Recall $U_t \equiv Y_t - X_t\beta(q)$ with $q = 0$ and $q = 1/2$ in the bounded and unbounded (lower) support cases respectively. In sequel, $F_t$ denotes $F_{u_t}(\cdot|X_t)$ and $E_x$ denotes expectation taken over $X_t$, etc.

$$\begin{aligned}EF_T^2 &= E_x|F_t[F_t^{-1}(\tau\underline{l})] - F_t[F_t^{-1}(\tau\underline{l}) + F_t^{-1}(\tau\bar{l}) - F_t^{-1}(\tau\underline{l}) + c/a_T]|c^2/\tau \\ &= E_x|f_t[F_t^{-1}(\tau\underline{l}) + w_{\tau t}] \times \{|F_t^{-1}(\tau\bar{l}) - F_t^{-1}(\tau\underline{l}) + c/a_T|\}|c^2/\tau\end{aligned}$$

by Taylor expansion, with $|w_{\tau t}|$ less than the term in $\{\cdot\}$ brackets. By the results of appendix for chapter 3 (Lemma A.4.1(v)) uniformly in $t$

$$F_t^{-1}(\underline{l}\tau) - F_t^{-1}(\bar{l}\tau) \sim C(x_t)[F_u^{-1}(m\tau) - F_u^{-1}(\tau)]] \tag{B.2.11}$$

where $0 < B_1 < C(x) < B_2$ for all $x$ and some constants $B_1$ and $B_2$. Since $c/a_T \equiv c[F_u^{-1}(m\tau) - F_u^{-1}(\tau)]/\sqrt{\tau T}$

$$|w_{\tau t}| \le m_{\tau t} \equiv |F_t^{-1}(\underline{l}\tau) - F_t^{-1}(\bar{l}\tau)|(1+o(1)) \quad \text{uniformly in } t. \tag{B.2.12}$$

By the assumed regular variation property, locally uniformly in $l$ and uniformly in $t$ :

$$f_t[F_t^{-1}(\tau l)] \sim l^{\xi+1} f_t[F_t^{-1}(\tau)].$$

That is, locally uniformly in $l, l'$ and uniformly in $t$ :

$$f_t[F_t^{-1}(\tau l') + F_t^{-1}(l\tau) - F_t^{-1}(l'\tau)] \sim (l/l')^{\xi+1} f_t[F_t^{-1}(\tau l')]. \tag{B.2.13}$$

By (B.2.12) it is possible to pick sequence $\{l_\tau, l'_\tau\} \to (\underline{l}, \bar{l})$ s.t. $m_{\tau t} = F_t^{-1}(l_\tau \tau) - F_t^{-1}(\bar{l}'_\tau \tau)$. Apply (B.2.13) to conclude that as $\tau \searrow 0$, uniformly in $t$

$$f_t[F_t^{-1}(\tau \underline{l}) + w_{\tau t}] \sim (\bar{l}/\underline{l})^{\xi+1} f_t[F_t^{-1}(\underline{l}\tau)]. \tag{B.2.14}$$

Therefore, for large $T$ and some $K > 0$, by (B.2.11), (B.2.14)and chapter 3,

$$\begin{aligned} EF^2 &\leq K(\bar{l}/\underline{l})^{\xi+1} f_t[F_t^{-1}(\underline{l}\tau)] \Big[ \Big| \frac{F_t^{-1}(\underline{l}\tau) - F_t^{-1}(\bar{l}\tau)}{\tau} \Big| + c \Big| \frac{F_u^{-1}(m\tau) - F_u^{-1}(\tau)}{\tau\sqrt{\tau T}} \Big| \Big] c^2 \\ &\leq K(\bar{l}/\underline{l})^{\xi+1} \Big[ \ \Big| \frac{(\bar{l}/\underline{l})^{-\xi} - 1}{-\xi} \Big| + O(\frac{1}{\sqrt{\tau T}}) \Big] = O(1). \end{aligned}$$

Also it is easy to see that $EF_T^2 1\{F_T > \epsilon\sqrt{T}\} = 0$ once $\tau T$ is sufficiently large. Hence property (i) is proved. To show property (ii) stated in the preceding proof, we first compute the bracketing number for class

$$\mathcal{F}'_T = \{1[X_1\beta(\tau l) \leq Y_1 \leq X_1(\beta(\tau l) + z/a_T)], (l, z) \in \mathcal{L} \times K\}.$$

Partition $\mathcal{L}$ into sets $[l_i, l_{i+1}]$ with $|l_i - l_{i+1}| \leq \delta$ and $\mathcal{K}$ – into cubes of side length $\delta$ with centers denoted as $z_j$. There are $O(1/\delta^d)$ such cubes. Define bracketing functions:

$$\begin{aligned} \bar{b}_k &= 1(x'\beta(\tau l_i) \leq y \leq x'\beta(\tau l_{i+1}) + (x'z_j + m\delta)/a_T)/\sqrt{\tau} \\ \underline{b}_k &= 1(x'\beta(\tau l_i) \leq y \leq x'\beta(\tau l_i) + x'z_j/a_T)/\sqrt{\tau} \end{aligned}$$

where $m = \sup_{x \in \mathbf{X}} \|x\|_2$. Calculations, much analogous to those above, yield that for some $K_1, K_2 > 0$ and $T$ sufficiently large

$$E\|\bar{b}_k - \underline{b}_k\|_2^2 \leq K_1 \Big| \frac{(l_{i+1}/l_i)^{-\xi} - 1}{-\xi} \Big| + O(\frac{1}{\sqrt{\tau T}}) \leq K_2\delta.$$

Thus $N_{[]}(\delta, \mathcal{F}'_T) = O(1/\delta^{2(d+1)})$, given construction of brackets. Finally, a trivial bound on bracketing number $N_{[]}(\delta, \mathcal{F}_T)$ is $O(1/\delta^{2(d+1)}) \times O(1/\delta^{2d})$, hence property (ii) holds [and thus $(\mathcal{L} \times K, \rho)$ is totally bounded]. Conclude as in the proof of part (i) of this lemma. ■

# Appendix C

# Appendix for Chapter 5

## C.1 Proofs

**Proof of Theorem 8.** Since (5.2.2) holds uniformly in $\Theta$ by Lemma A.4.1, it suffices to show that pointwise or uniformly in $\Theta$ the empirical regression quantile spacings can consistenly replace the population quantile spacings:

$$\frac{x'(\hat{\beta}(mls\tau) - \hat{\beta}(ls\tau))}{x'(\beta(mls\tau) - \beta(ls\tau))} \xrightarrow{p} 1.$$

We give the proof of the uniform result only. Proof of the pointwise result is similar. First, uniformly in $\theta$, by Lemma A.4.1:

$$\begin{aligned}\frac{x'(\hat{\beta}(mls\tau) - \hat{\beta}(ls\tau))}{x'(\beta(mls\tau) - \beta(ls\tau))} &= \frac{x'(\hat{\beta}(mls\tau) - \hat{\beta}(ls\tau))}{\mu_X'(\beta(mls\tau) - \beta(ls\tau))} \frac{\mu_X'(\beta(mls\tau) - \beta(ls\tau))}{x'(\beta(mls\tau) - \beta(ls\tau))} \\ &\sim \frac{x'(\hat{\beta}(mls\tau) - \hat{\beta}(ls\tau))}{\mu_X'(\beta(mls\tau) - \beta(ls\tau))} \frac{1}{\mathcal{H}(x)}\end{aligned}$$

where $\mathcal{H}(x) = x'\mathbf{c}$ for type 2 and 3 tails, and $\mathcal{H}(x) = 1 -$ for type 1 tails. $x'\mathbf{c}$ is positive and bounded away from zero and above on $\mathbf{X}$. Define $A_\theta = \frac{x'(\hat{\beta}(mls\tau) - \beta(mls\tau))}{\mu_X'(\beta(mls\tau) - \beta(ls\tau))}$, $B_\theta = \frac{x'(\beta(ls\tau) - \hat{\beta}(ls\tau))}{\mu_X'(\beta(mls\tau) - \beta(ls\tau))}$, $C_\theta = \frac{x'(\beta(mls\tau) - \beta(ls\tau))}{\mu_X'(\beta(mls\tau) - \beta(ls\tau))}$. Then

$$\frac{x'(\hat{\beta}(mls\tau) - \hat{\beta}(ls\tau))}{\mu_X'(\beta(mls\tau) - \beta(ls\tau))} \frac{1}{\mathcal{H}(x)} = (A_\theta + B_\theta + C_\theta)\frac{1}{\mathcal{H}(x)}$$

By Theorem 7, $A_\theta, B_\theta = O_{p*}(1/\sqrt{\tau T}) = o_{p*}(1)$ uniformly in $\theta$ in $\Theta$. By Lemma A4.1, $C_\theta \sim \mathcal{H}(x)$, uniformly in $\theta \in \Theta$. ■

**Proof of Theorem 9.** Proof is very similar to the proof of Theorem 8.■

## C.2 Monte-Carlo Illustrations

A first question to ask is whether the asymptotic theory developed here makes any sense at all. One may question both its value and even correctness. An extensive Monte-Carlo work is contained in [19], supporting the theory. Our goal here is to briefly illustrate whether our asymptotic theory is a sensible guide in moderate and even small samples and whether predictions of the theory work as the sample size increases. It is also useful to discuss the usefulness and limitations of the theory from the practical inference view point.

We first illustrate the performance of the asymptotic distributions under the extreme rank conditions. The essence of simulation work in [19] is perhaps best conveyed through an example:

- $Y_t = X_t'1 + U_t$,
- $\{U_t\}$ has type 2 tails (Cauchy), i.i.d., independent of $\{X_t\}$.
- $X_t$ are i.i.d., representative of common regressors used in econometrics.

Simulation design is as follows. $X_1 \sim Beta(3,3)$ has a symmetric distribution, $X_2 \sim Beta(1.5,6)$ has a skewed distribution. $X_2$ is independent of $X_1$. $X_3$ and $X_4$ are categorical "dummy" variables taking values 1 or 0. $P(X_1 = 1) = 1/3$, $P(X_4 = 1) = 1/3$, $P(X_4 = 1, X_3 = 1) = 0$, $P(X_4 = 0, X_3 = 0) = 1/3$. The design is thus representative of regressors in economic applications: symmetrically and asymmetrically distributed and categorical variables.

Suppose regression quantiles of interest have rank k=1 and k=5. At first, set the sample size small: $T = 50$. Figure C.2 depicts the true (simulated) distribution of the estimators $\hat{\beta}(\tau) - \beta(\tau)$ for $\tau = 1/T$ (solid line). The dashed line is the approximating extreme rank density function. It is also simulated. Number of replications is 1,000. . The plots correspond to $\hat{\beta}_1(\tau) - \beta_1(\tau)$, ..., $\hat{\beta}_d(\tau) - \beta_d(\tau)$. The lines are fairly close but clearly distinguishable. (More formal comparisons, such as coverage levels, appear in [19]. We prefer the graphical method, for it, in our opinion, conveys all the information without loss and need to justify one or the other numeric criterion). The flat line with circles is the regular (central rank) normal approximation, depicted for reference. Next we set $k = 5$ (figure C.3). The true and the extreme rank distributions are close but clearly less so than in the case with $k = 1$. The normal approximation looks here more reasonable than for the case with $k = 1$. The extreme rank approximation is able to reproduce the skewed shape of distribution of

$\hat{\beta}_1(\tau)-\beta_1(\tau)$, the intercept coefficient. It also matches well the distributions of $\hat{\beta}_3(\tau)-\beta_3(\tau)$ and $\hat{\beta}_4(\tau)-\beta_4(\tau)$ (associated with the categorical variables). It is also able to reproduce the kurtosis features of the true distributions. Although the regular normal approximation clearly does not match such features, it does not appear to be unreasonable. One may argue that, for inference purposes, it may well be a sensible (within certain limits) distribution to use. However, one should be careful here in view of unknown nuisance parameters. (The relevant nuisance parameter in the regular theory is $1/f_u(F_u^{-1}(\tau))$, that is the density function. Now imagine using conventional non-parametric methods to estimate this density with only few observations in the tail.) We shall come back to this issue. For now we would like to separate the issues of the distribution theory from the nuisance parameter problem.

Does the extreme rank theory work as we increase the sample size? It clearly does. Let us increase the sample size up to 1000 and keep ranks $k = 1$ and $k = 5$ fixed (as the theory requires). Figures C.4 and C.5 describe the changes. Now the lines of the true and the extreme rank density are fairly close and sometime very close. Normal approximation looks somewhat sensible for $k = 5$. Thus the two important questions are answered. Many other examples with various distributional assumptions and sample sizes are contained in [19]. The results appear to be robust to these further variations.

The next step is to compare the intermediate and the central rank theories. Both theories yield the limiting normal distributions, albeit different variance-covariance matrices and scalings. Therefore, intermediate rank approximation can not compete with the extreme rank approximations when rank $k$ is low, as it will not be able to reproduce the kurtosis and skewness of the distributions we saw. When can it compete with the regular normal? Clearly, only when $\tau$ is low and $\tau T$ is rather large. Under such conditions the two approximations should give similar answers (after all, the intermediate rank approximation may be viewed informally as a limit of the regular distribution). The value of the intermediate rank approximation is realized in practical work, when nuisance parameters must be estimated. The discussion is yet postponed until the two main questions are answered.

Again let us consider the same setting as before. First we start with $T = 100$ and rank $k = 15$ ( $\tau = .15$). Figure C.6 indicates the quality of different approximations in small sample. The scaling $a_T$, in the intermediate rank approximation, that depends on an arbitrary fixed constant $m > 0$. We used $m = 1.5$ (in the unconditional analysis it is common to choose $m = 2$, see [31]). Both central (denoted by line with circles) and intermediate rank approximations (dashed line) are close. They both perform moderately well as approximations of the true distributions (solid line). Thus small sample performance is reasonable. For reference we also depicted the extreme rank distribution (lines with letter "e"). Surprisingly, it does quite well as an approximation. Note that none of the approximations are able to match the bias of the true distribution of the intercept regression quantile coefficient. To eliminate the bias, one needs to have a sufficiently high $k$. $k = 15$ does not appear to satisfy this requirement. The intermediate asymptotics requires rank to grow and quantile index to converge to zero. Select now the sample of $T = 1000$. Select $k = 40 > 15$ $(\tau = .04)$ . The distributions are now much closer to the true distribution and bias is much less (figure C.7).

These results suggest that we have obtained plausible approximations of the true distributions of regression quantile coefficients. These notions exploit the asymptotically formulated data scarcity conditions. Extreme value approximation performs quite well in all situations, even for significantly high rank. Both intermediate and central approximation perform well when the rank is sufficiently high. But more serious issue is the performance in cases of unknown nuisance parameters. The informational requirements of all methods are different. Extreme ranks require an estimate of the scaling constants, which often can be estimated only up to a constant, but in such cases, fortunately, on can resort to subsampling in order to construct confidence intervals. Under simplest assumptions (such as exact algebraic tails, for example) we only need the estimate of the tail index to carry out subsampling. More complicated cases with unknown scaling constants are discussed in Bertail et.al., where subsampling itself is used to estimate the unknown rate of convergence. These studies indicate a plausible performance of subsampling under rather general conditions. Thus we could successfully exploit this quality in order to construct confidence intervals in our problem. (In appendix C, it is shown how certain centering constants validate subsampling together with the result on the existence of the limit law).

Central ranks require the density estimate in the tails. This is a hard problem. Current packages incorporating regression quantiles use the methods that are adequate for the center of the distribution. Consider next the intermediate rank approximation. Given that we can estimate the scaling constant consistently as demonstrated in Chapter 3, the only remaining nuisance *parameter* is the tail index (and the tail heterogeneity index), which is easily estimable in a (semi)-parametric fashion, using the L-statistics of the ratios of spacings of regression quantiles. Our Monte-Carlo work[19] indicates good performance of the estimators of this kind. Furthermore, the resulting estimate of the true variance of regression coefficients is significantly better and more robust than the standard centrality based estimators. Hence despite that both intermediate and central rank approximations are similar, the value of intermediate ranks is revealed through the ease and quality of inference. See [19] for further discussion.

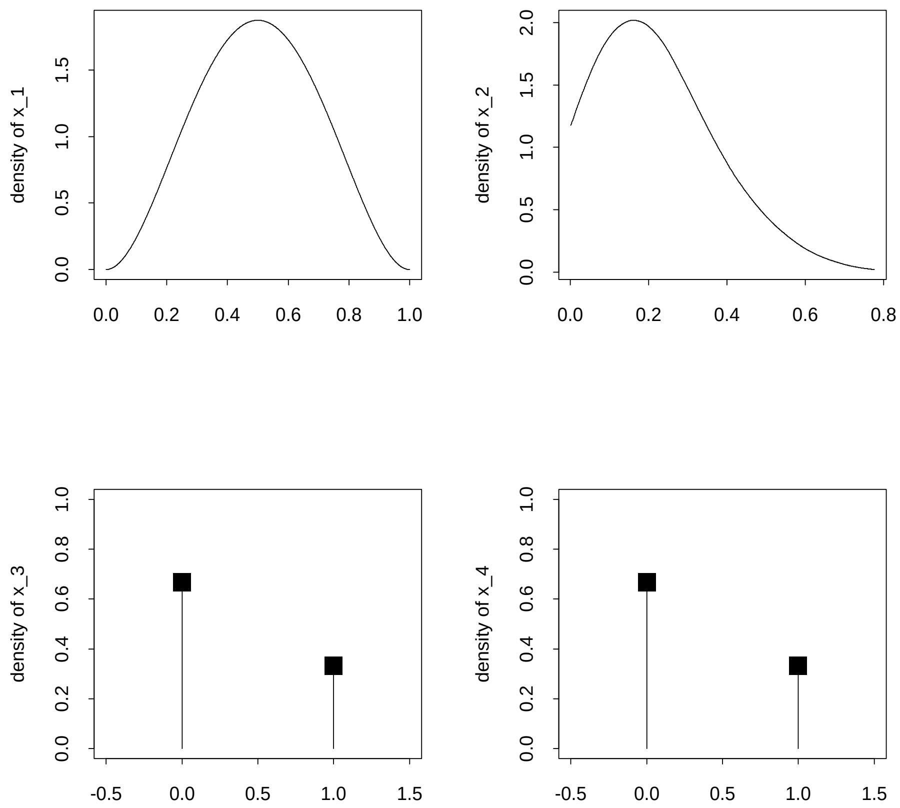


Figure C.1: Simulation Design: Marginal Densities of Regressors.

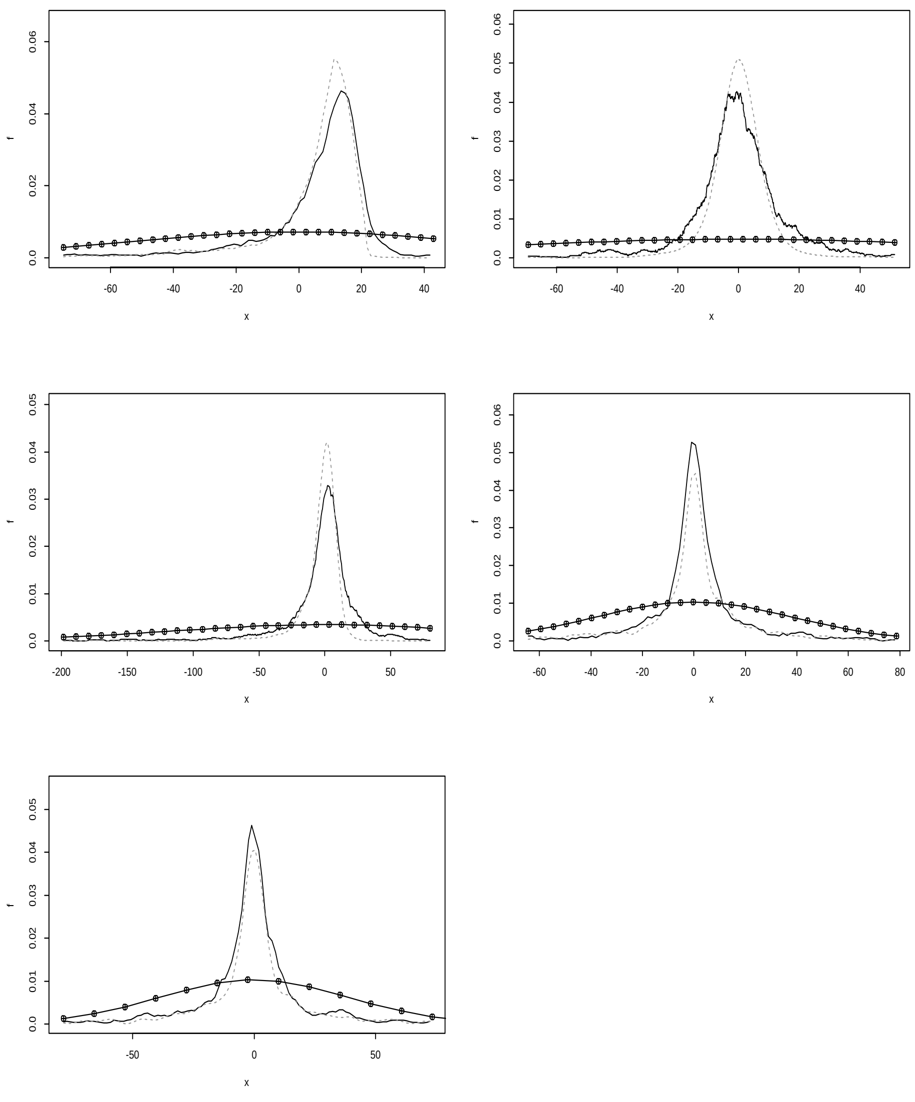

Figure C.2: Small Sample Approximation Properties. Case when $k = 1$, $T = 50$

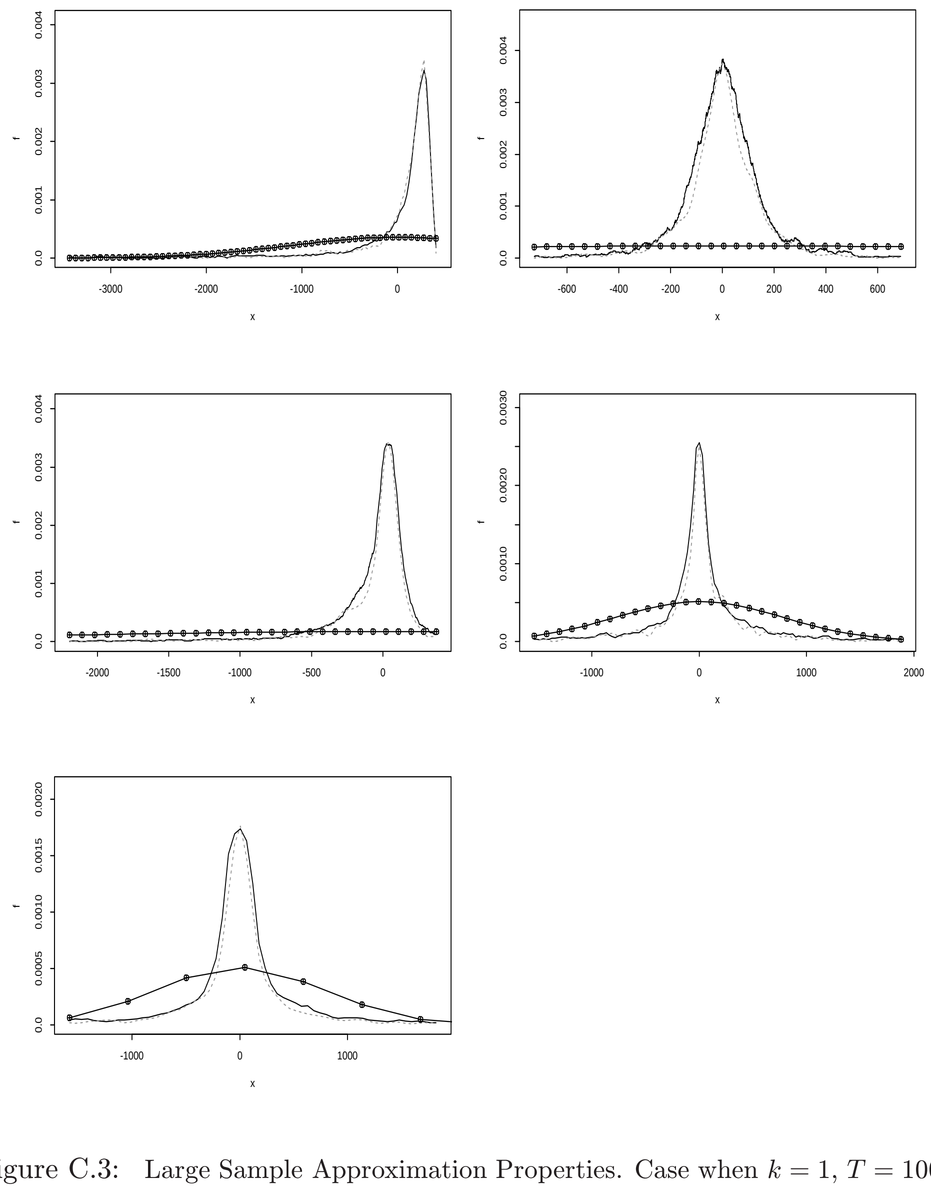

Figure C.3: Large Sample Approximation Properties. Case when $k = 1$, $T = 1000$.

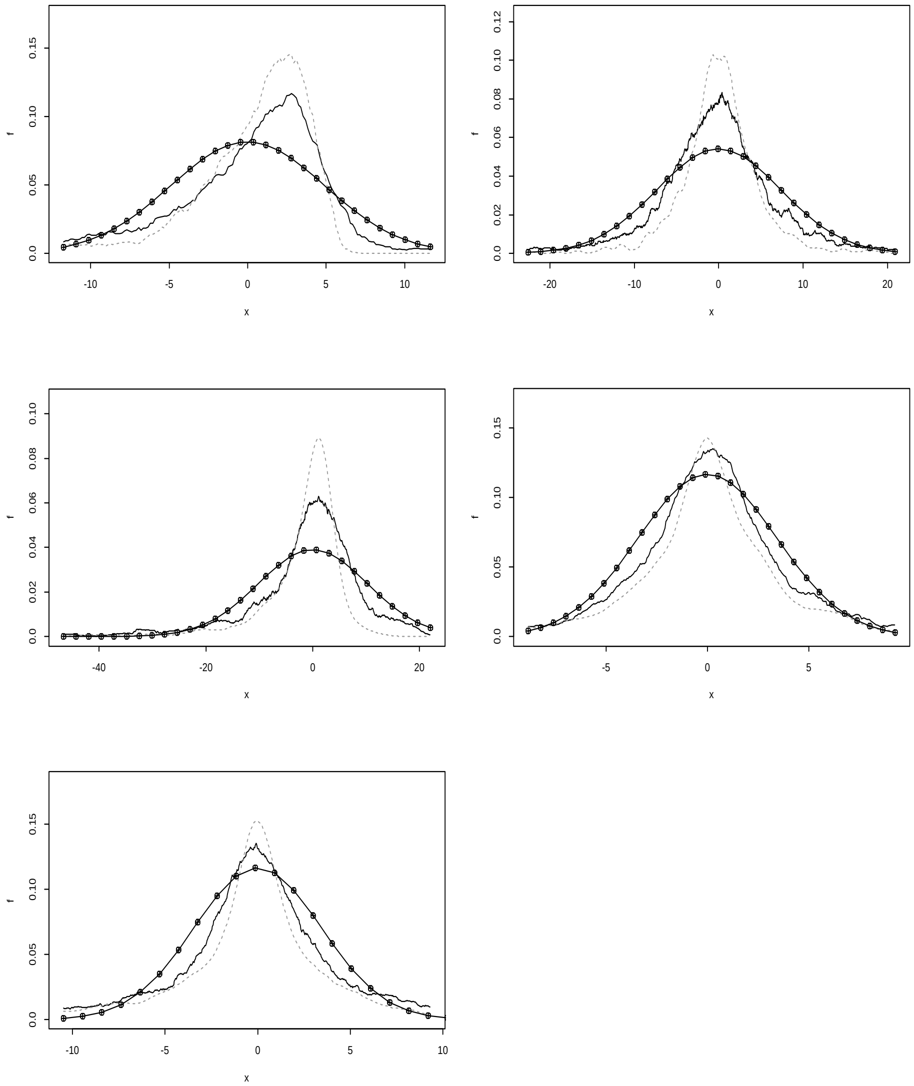


Figure C.4: Small Sample Approximation Properties. Case when $k = 5$, $T = 50$.

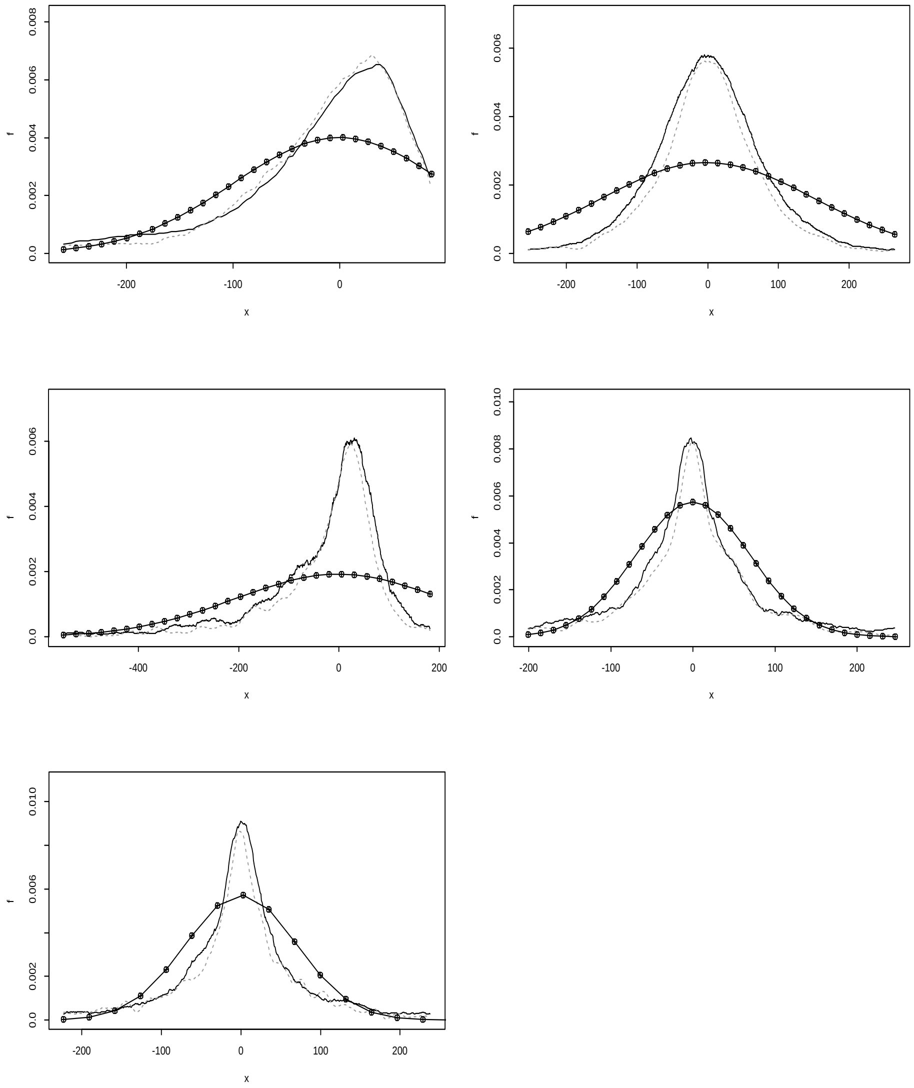


Figure C.5: Large Sample Approximation Properties. Case when $k = 5$, $T = 1000$.

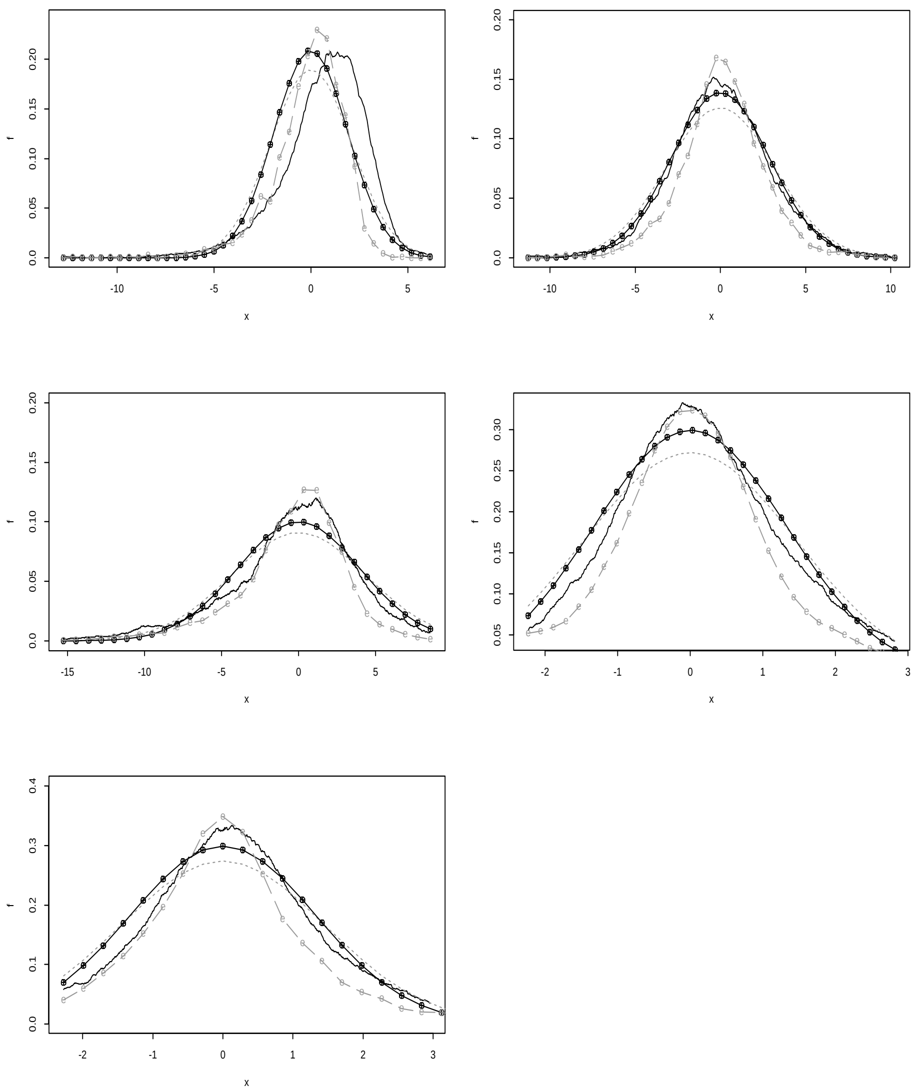


Figure C.6: Small Sample Approximation Properties. Case when $k = 15$, $T = 100$.

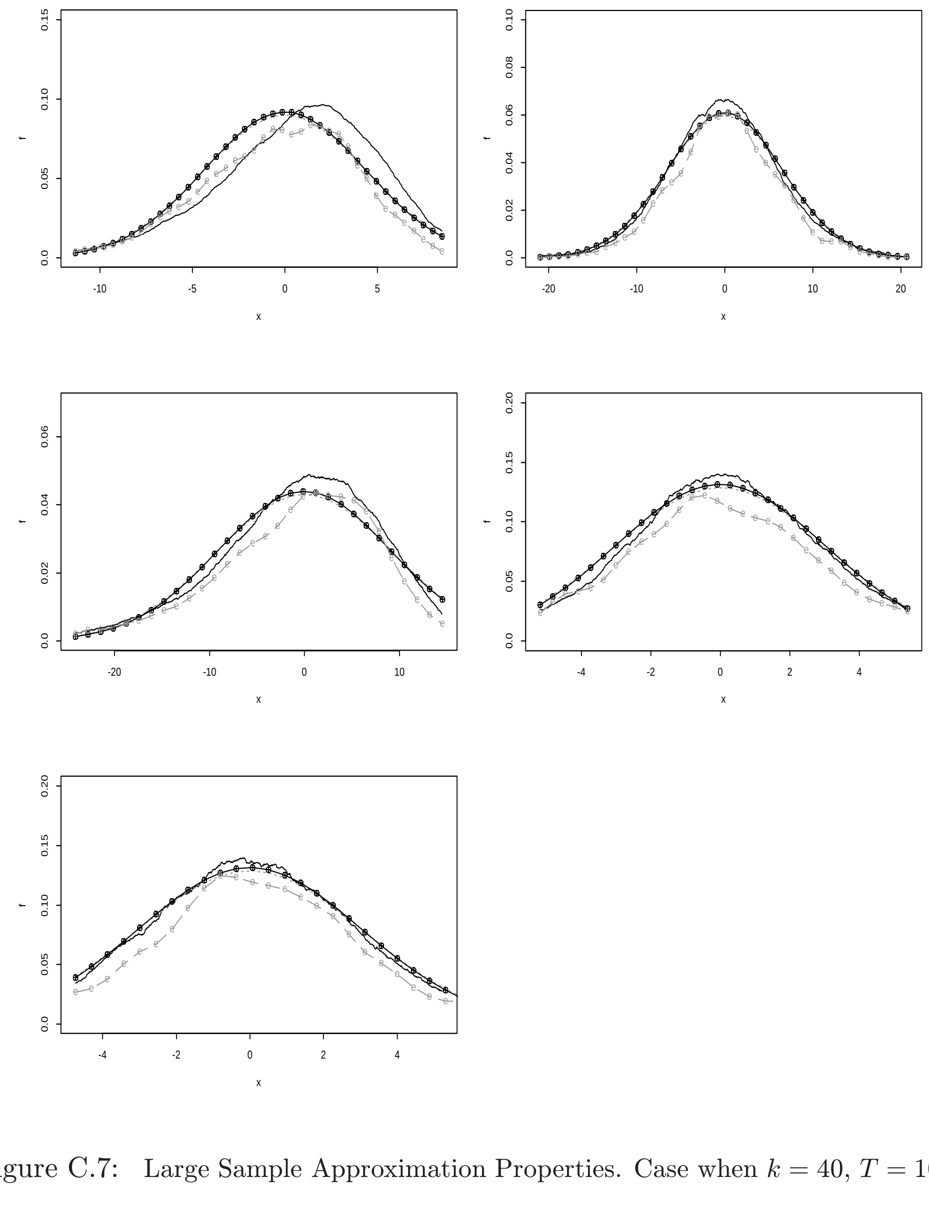

Figure C.7: Large Sample Approximation Properties. Case when $k = 40$, $T = 1000$.

## C.3 Estimating Unknown Variance: Comparison of Methods

**Definitions and Calculations of Criteria**

MSE is the relative mean squared error:

$$Mse(\widehat{var}(\hat{\beta}_i(\tau)) - var(\hat{\beta}_i(\tau)))/var(\hat{\beta}_i(\tau))^2$$

MAD is the relative mean absolute deviation:

$$Mad(\widehat{var}(\hat{\beta}_i(\tau)) - var(\hat{\beta}_i(\tau)))/var(\hat{\beta}_i(\tau))$$

MeDAD is the median relative absolute deviation:

$$Median(|\widehat{var}(\hat{b}eta_i(\tau)) - var(\hat{\beta}_i(\tau))|)/var(\hat{\beta}_i(\tau))$$

For each of the regression coefficients, we computed the true variance $var(\hat{\beta}_i(\tau))$, $i = 1, .., 5$. For each replication we computed $\widehat{var}(\hat{\beta}_i(\tau))$ by the intermediate rank method and by the central rank methods (Bofinger's bandwidth choice, implemented in S-Plus). We then computed the above measures for the intermediate and central methods. Number of replications is 4,000. The following tables present some key results. More can be found in [19].

| relative criterion* | $i = 1$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| mse interm | 1.40945 | 1.35737 | 1.36146 | 1.46699 | 1.3019 |
| mse central | 81.0908 | 77.5948 | 77.8708 | 84.9085 | 73.8221 |
| mad interm | 0.788007 | 0.781195 | 0.781734 | 0.795336 | 0.773733 |
| mad central | 8.90183 | 8.70738 | 8.72289 | 9.10944 | 8.49255 |
| medad interm | 0.657331 | 0.66045 | 0.659953 | 0.659715 | 0.647428 |
| medad central | 8.8615 | 8.66784 | 8.68328 | 9.06826 | 8.45389 |

Table C.1: Performance of Regular Methods (Central Rank) Relative to Intermediate Rank Methods of Estimating True Variance of Quantile Regression Coefficients. $k = 5(\tau = k/T)$, $T = 200$, Type 1 Errors (Normal). Regressor design is described in a previous figure.

| relative criterion* | $i = 1$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| mse interm | 0.576481 | 0.54829 | 0.539976 | 0.677342 | 0.622835 |
| mse central | 73.3936 | 68.7659 | 67.3517 | 88.4579 | 80.5566 |
| mad interm | 0.57403 | 0.567706 | 0.565844 | 0.597762 | 0.584709 |
| mad central | 8.46781 | 8.19584 | 8.11091 | 9.29832 | 8.87239 |
| medad interm | 0.501793 | 0.5075 | 0.510497 | 0.495884 | 0.493222 |
| medad central | 8.42925 | 8.15838 | 8.0738 | 9.25637 | 8.83218 |

Table C.2: Performance of Regular Methods (Central Rank) Relative to Intermediate Rank Methods of Estimating True Variance of Quantile Regression Coefficients. $k = 10(\tau = k/T)$, $T = 200$, Type 1 Errors (Normal). Regressor design is described in a previous figure.

| relative criterion* | $i = 1$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| mse interm | 0.337869 | 0.324067 | 0.34022 | 0.433336 | 0.389063 |
| mse central | 57.1852 | 53.6817 | 57.7498 | 76.237 | 68.1494 |
| mad interm | 0.460687 | 0.456782 | 0.461377 | 0.494608 | 0.477943 |
| mad central | 7.47209 | 7.23893 | 7.50899 | 8.63068 | 8.15893 |
| medad interm | 0.41893 | 0.413758 | 0.417854 | 0.421953 | 0.413184 |
| medad central | 7.43758 | 7.20538 | 7.47433 | 8.59145 | 8.12162 |

Table C.3: Performance of Regular Methods (Central Rank) Relative to Intermediate Rank Methods of Estimating True Variance of Quantile Regression Coefficients. $k = 15(\tau = k/T)$, $T = 200$, Type 1 Errors (Normal). Regressor design is described in a previous figure.

# Bibliography


[1] Ait-Sahalia, Y., and Lo, A. W. Nonparametric risk management and implied risk aversion. *preprint* (1998).

[2] Amemiya, T. *Advanced Econometrics*. Harvard University Press, 1985.

[3] Anderson, E., Hansen, L. P., and Sargent, T. J. Risk and robustness in equilibrium. Manuscript, 1999.

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

[5] Bajari, P. Auction models are not robust when bidders make small mistakes. prperint, Stanford, 1999.

[6] Bassett, Jr., G., and Koenker, R. An empirical quantile function for linear models with iid errors. *J. Amer. Statist. Assoc. 77*, 378 (1982), 407–415.

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

[8] Bierens, H. Kernel estimators of regression functions. In *Advances in economic theory*, T. Bewley, Ed. Cambridge University Press, 1989. the Fifth World Congress of the Econometric Society, August 1985.

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

[10] Caballero, R. Aggregate investment. *NBER Working Paper Series 6264* (1998).

[11] Caballero, R., and Engel, E. Dynamic (S,s) economies. *NBER Working Paper Series 3734* (1991).

[12] Caballero, R., and Engel, E. Nonlinear aggregate investment dynamics: Theory and evidence. *NBER Working Paper Series 6420* (1998).

[13] Caballero, R., and Engel, E. Explaining investment dynamics in U.S. manufacturing: a generalized (S,s) approach. *Econometrica 67*, 4 (1999), 783–826.

[14] Caballero, R., and Leahy, J. V. Fixed costs: The demise of marginal q. *NBER Working Paper Series 5508* (1997).

[15] Caplin, A., and Leahy, J. Durable Goods Cycles. NBER Working Paper No. 6987, 1999.

[16] Chamberlain, G. Quantile regression censoring and the structure of wages. In *Advances in Econometrics*, C. Sims, Ed. Elsevier: New York, 1994.

[17] Chernozhukov, V. Nonparametric extreme regression quantiles. mimeo, 1998.

[18] Chernozhukov, V. Inference for extreme and near-extreme regression quantiles. in progress , Stanford, 1999.

[19] Chernozhukov, V. Inference for extremal quantile regression: prospects and simulation study. Stanford., 2000.

[20] Chernozhukov, V., and Umantsev, L. Regression VaR: Modeling and Estimation. August, mimeo, Stanford University, 1999.

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

[22] Čibisov, D. M. Limit distributions for order statistics. *Teor. Verojatnost. i Primenen. 9* (1964), 159–165.

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

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

[25] Davis, R. A., Knight, K., and Liu, J. $M$-estimation for autoregressions with infinite variance. *Stochastic Process. Appl. 40*, 1 (1992), 145–180.

[26] Dekkers, A. L. M., and de Haan, L. On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist. 17*, 4 (1989), 1795–1832.

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

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

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

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

[31] Embrechts, P., Klüppelberg, C., and Milkosh, T. *Modelling Extremal Events*. Applications of Mathematics. Stochastic Modelling and Applied Probability. Springer, 1997.

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

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

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

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

[36] Galombos, J. *The Asymptotic Theory of Extreme Order Statistics*. Wiley, Ney York, 1978.

[37] Galombos, J. *Order Statistics. In:Handbook of Statistics. Vol4, Eds. P.R.Krishnaiah and P.K.Sen, pp. 359-382.* Amsterdam:North-Holland, 1984.

[38] HALL, P., AND HEYDE, C. C. *Martingale limit theory and its application.* Academic Press, New York, 1980.

[39] HECKMAN, J., AND MACURDY, T. Labor econometrics. *Handbook of Econometrics 3* (1986).

[40] HILL, B. M. A simple general approach to inference about the tail of a distribution. *Ann. Statist. 3*, 5 (1975), 1163–1174.

[41] HOFFMANN-JØRGENSEN, J. *Stochastic processes on Polish spaces.* Aarhus Universitet, Matematisk Institut, Aarhus, 1991.

[42] HOFFMANN-JØRGENSEN, J. *Probability with a view toward statistics.* Chapman & Hall, New York, 1994.

[43] HONG, H. Non-regular maximum likelihood estimation in auction, job search and production frontier models. mimeo., 1997.

[44] HONG, H., AND SHUM, M. A Structural Econometric Ascending Auction Model, with an Application to the PCS Spectrum Auctions. Mimeo, 1998.

[45] JUREČKOVÁ, J. Asymptotic relations of $M$-estimates and $R$-estimates in linear regression model. *Ann. Statist. 5*, 3 (1977), 464–472.

[46] KALLENBERG, O. *Random measures*, fourth ed. Akademie-Verlag, Berlin, 1986.

[47] KERNES, P., AND PAGAN, A. Estimating the density tail index for financial time series. *Review of Economics and Statistics 79*, 2 (1997), 171–175.

[48] KNIGHT, K. Epi-convergence and stochastic equisemicontinuity. Preprint, 1999.

[49] KNIGHT, K. Limiting distributions of the linear programming estimator. Preprint, 1999.

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

[51] KOENKER, R., AND D'OREY, V. Computing regression quantiles. *Applied Statistics 36* (1987), 383–390.

[52] Koenker, R., and Machado, J. Goodness of Fit and Related Inference Processes for Quantile Regression. preprint, 1999.

[53] Koenker, R., and Park, B. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics 71* (1996), 265–283.

[54] Koenker, R., and Portnoy, S. The Gaussian Hare and the Laplacian Tortoise. *Statistical Science 12* (1997), 279–300.

[55] Koenker, R., and Portnoy, S. Regression quantiles. preprint, 1999.

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

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

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

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

[60] Mosteller, F. On some useful "inefficient" statistics. *Ann. Math. Statistics 17* (1946), 377–408.

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

[62] Politis, D. N., Romano, J. P., and Wolf. Subsampling (book in preparation). *Stats Dept, Stanford* (1999).

[63] Pollard, D. *Convergence of stochastic processes.* Springer-Verlag, New York-Berlin, 1984.

[64] Pollard, D. *Empirical Processes: Theory and Applications.* Hayard, CA: Institute of Mathematical Statistics, 1990.

[65] Pollard, D. Asymptotics for least absolute deviation regression estimator. *Econometric Theory 7* (1991), 186–199.

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

[67] Portnoy, S. Asymptotic behavior of the number of regression quantile breakpoints. *SIAM J. Science Statistical Computing 12* (1989), 867–883.

[68] Portnoy, S., and Jureckova, J. On extreme regression quantiles. preprint, 1998.

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

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

[71] Resnick, S. *Extreme Values, Regular Variation, and Point Processes.* Springer, 1987.

[72] Resnick, S. I. *Extreme values, regular variation, and point processes.* Springer-Verlag, New York-Berlin, 1987.

[73] Robinson, P. M. Nonparametric estimators for time series. *J. Time Ser. Anal. 4*, 3 (1983), 185–207.

[74] Rockafellar, R. T., and Wets, R. B. *Variational Analysis.* Springer-Verlag, Berlin, 1998.

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

[76] Salinetti, G., and Wets, R. B. 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.

[77] Sargent, T. J. *Dynamic Macroeconomic Theory.* Harvard University Press, 1989.

[78] Schwarz, M. Decision-Making under Extreme Uncertainty. Stanford University, job market paper, 1998.

[79] Smirnov, N. V. Convergence of distributions of order statistics to the normal distribution. *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk 10*, 3 (1966), 24–32.

[80] Smith, R. L. Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika 72* (1985), 67–90.

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

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

[83] Umantsev, L., and Chernozhukov, V. Risk: Polynomial quantile models. September, mimeo, Stanford University, 1999.

[84] van der Vaart, A. W. *Asymptotic statistics.* Cambridge University Press, Cambridge, 1998.

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

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

[87] Watts, V., Rootzén, H., and Leadbetter, M. R. On limiting distributions of intermediate order statistics from stationary sequences. *Ann. Probab. 10*, 3 (1982), 653–662.

[88] Weitzman, M. L. Optimal search for the best alternative. *Econometrica 47* (1979), 641–654.