Asymptotic Inference for the Constrained Quantile Regression Process

Thomas Parker
Department of Economics, University of Waterloo

December 20, 2018

Abstract

I investigate the asymptotic distribution of linear quantile regression coefficient estimates when the parameter lies on the boundary of the parameter space. In order to allow for inferences made across many conditional quantiles, I provide a uniform characterization of constrained quantile regression estimates as a stochastic process over an interval of quantile levels. To do this I pose the process of estimates as solutions to a parameterized family of constrained optimization problems, parameterized by quantile level. A uniform characterization of the dual solution to these problems — the so-called regression rankscore process — is also derived, which can be used for score-type inference in quantile regression. The asymptotic behavior of quasi-likelihood ratio, Wald and regression rankscore processes for inference when the null hypothesis asserts that the parameters lie on a boundary follows from the features of the constrained solutions.

Keywords: Quantile regression, inequality constraints, asymptotic inference
JEL Classification Codes: C12, C21, C31

1 Introduction

Quantile regression is a standard tool for investigating the relationship between covariates and the distribution of a response variable. Some research has investigated quantile regression estimation when coefficients are subject to inequality constraints. For example, penalized spline estimation for conditional quantile functions subject to qualitative constraint such as monotonicity has been developed in Koenker and Ng (2005), and estimation and bootstrap inference for isotonic quantile regression was investigated in Abrevaya (2005). Furthermore, the general methods developed in Andrews (2001) apply to inference for quantile regression coefficients made at a single quantile level. However, asymptotic inference for inequality constrained quantile regression estimates that is suited to simultaneous inference over several quantile levels has not been investigated. In this paper I provide an asymptotic characterization of linear quantile regression estimates as a stochastic process in quantile level — thereby addressing simultaneous inference over several quantiles — when the estimates are made under the imposition of inequality constraints.

Asymptotics for the normal location problem under constraints have been very well studied, and a detailed exposition is available in a series of well known monographs (Barlow et al., 1972; Robertson et al., 1988; Silvapulle and Sen, 2005). Asymptotic properties of the quantile regression estimator for a single quantile level and other M-estimators under constraints are discussed in Geyer (1994) and Andrews (1999). Andrews (2001) provides a very general approach to tests based on these estimates. However, quantile regression has a rather special feature among M-estimators: an analyst may be concerned with simultaneous inference for multiple (even an interval of) quantile levels. In order to apply to hypotheses on all relevant configurations of quantile levels, the quantile regression objective function and its solution are considered as stochastic processes over a set of quantile levels $\mathcal{T} = [\epsilon, 1-\epsilon]$ for some $\epsilon \in (0, 1/2)$. This mirrors the inference processes developed in Koenker and Machado (1999) except that here estimates are made under the imposition of inequality constraints.

In the next section some technical tools are established that are necessary for dealing with the asymptotics of constrained quantile regression objective function processes and estimate processes. In Section 3 the dual solution of the constrained coefficient estimation problem is introduced, which is called the constrained regression rankscore process, and is shown to provide a uniform approximation to the subgradient of the objective function process. In Section 4 asymptotic theory for the distribution of constrained estimates is used to propose three inference processes that serve as generalizations of likelihood ratio, Wald and score or Lagrange multiplier tests of linear restrictions to tests over quantile levels contained in the index set $\mathcal{T}$. An appendix contains information on constrained parameter subvectors, some technical results about regression rankscores and two small simulation experiments illustrating the behavior of supremum- and L^2-norm test statistics based on the inference processes developed in Section 4.

2 Asymptotic theory for constrained coefficient estimates

Consider a linear model for the conditional quantile function of the response variable $Y \in \mathbb{R}$ given covariates $X \in \mathbb{R}^p$:

$$Q_{Y|X}(\tau|X = x) = x^\top \beta(\tau) \tag{1}$$

where $\beta(\tau) \in \mathbb{R}^p$. Assuming this is a reasonable description of the data, the linear quantile regression estimator can capture features of the conditional distribution of Y given X in a parsimonious fashion. Suppose that a researcher maintains the hypothesis that $\beta(\tau)$ lies in some subset of $\mathbb{R}^p$ and estimates coefficients under this hypothesis. For example, estimation techniques developed in Koenker and Ng (2005) can be used to estimate quantile regressions under the inequality-constrained hypothesis $R\beta(\tau) \geq r$ with $R \in \mathbb{R}^{q \times p}$, $q \leq p$ and $r \in \mathbb{R}^q$. Inference may be conducted under the maintained hypothesis $K : R\beta(\tau) \geq r$, testing the null and alternative hypotheses $H_0 : R\beta(\tau) = r$ vs. $H_1 : R\beta(\tau) > r$. Alternatively the maintained hypothesis may be $K : R\beta(\tau) \in \mathbb{R}^p$ with null and alternative hypotheses $H_0 : R\beta(\tau) \geq r$ vs. $H_1 : R\beta(\tau) \ngeq r$.

The quantile function of a distribution is the solution to a convex optimization problem. Specifically, suppose that Y is a univariate random variable with quantile function Q and define the objective function

$\rho_\tau(u) = u(\tau - I(u < 0))$, where $I(A) = 1$ when event A is true and is zero otherwise. Then $Q(\tau) = \text{argmin}_m \mathrm{E}[\rho_\tau(X-m)]$ (Koenker, 2005, Section 1.3). Given a sample of iid realizations of the random variable $\{Y_i\}_{i=1}^n$, the τ^{th} sample quantile can be estimated by computing $\hat{Q}(\tau) = \text{argmin}_m \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - m)$. Quantile regression makes the leap from this univariate model to a regression model by generalizing the location parameter in the univariate optimization problem to a linear function.

Given a sample $\{Y_i, X_i\}_{i=1}^n$, the τ^{th} linear quantile regression coefficient estimate is defined as

$$\hat{\beta}(\tau) = \operatorname*{argmin}_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau\left(Y_i - X_i^\top b\right). \tag{2}$$

By varying τ one has a convenient summary of the way that covariates affect different regions of the conditional distribution of Y, since $\beta_j(\tau) = \frac{\partial}{\partial X_j} Q_{Y|X}(\tau|X)$. The solution path over τ also represents the set of solutions to a parametric family of optimization problems where the parameter is $\tau \in \mathcal{T}$.

A substantial literature has described inference methods for the solution to (2), either at specific quantile levels or uniformly over many quantiles. Call $\hat{\beta}$ an unconstrained estimate of β when it is a minimizer in $\mathbb{R}^p$, as in (2). The *constrained* quantile regression estimator $\tilde{\beta}(\tau)$ for any $\tau \in \mathcal{T}$ is defined as

$$\tilde{\beta}(\tau) = \operatorname*{argmin}_{b \in \mathcal{B}(\tau)} \sum_{i=1}^n \rho_\tau\left(Y_i - X_i^\top b\right)$$

for a $\mathcal{B}(\tau) \subset \mathbb{R}^p$ such as $\mathcal{B}(\tau) = \{\beta : R\beta \geq r\}$. One may be interested in inference for several quantiles at once — for example, whether $\frac{\partial}{\partial X_j} Q_{Y|X}(\tau_k|X) > 0$ over several $\{\tau_k\}$, where an analyst imposes the maintained hypothesis that $\tilde{\beta}_j(\tau_k) \geq 0$ for all k in estimation. For a complete asymptotic description it is most appropriate to develop inference methods for $\tilde{\beta}$ as a stochastic process in which intervals of quantile levels $\tau \in \mathcal{T}$ may be considered. The main challenge in considering the constrained quantile regression process is that given a finite sample the set of binding constraints is random and depends on the observed sample and the desired quantile level.

The epigraph of the quantile regression objective function is a tractable starting point for the analysis of constrained quantile regression solutions. Recall that the epigraph of a function $f : \mathbb{R}^k \to \mathbb{R}$ is the set $\text{epi}f := \left\{(x, \alpha) \in \mathbb{R}^{k+1} : \alpha \geq f(x)\right\}$. For any $k \geq 1$ let $\mathcal{F}(\mathbb{R}^k)$ be the collection of all closed sets in $\mathbb{R}^k$. Functions are identified with their epigraphs and thereby asymptotic properties depend on the convergence of a sequence of epigraphs, which are elements of $\mathcal{F}(\mathbb{R}^{k+1})$. A convenient metric for measuring distance is the Attouch-Wets metric $d_{\mathcal{F}} : \mathcal{F}(\mathbb{R}^k) \times \mathcal{F}(\mathbb{R}^k) \to \mathbb{R}$ defined here by

$$d_{\mathcal{F}}(A, B) = \sum_{j=1}^\infty \frac{1}{2^j} \sup_{d(x,0)<j} |d(x, A) - d(x, B)| \wedge 1,$$

where $d : \mathbb{R}^k \times \mathcal{F}(\mathbb{R}^k) \to \mathbb{R}$ is the Euclidean distance function $d(x, C) = \min_{y \in C} \|x - y\|$ (Beer, 1993, p. 79). This metric makes it possible to discuss the convergence of one epigraph to another, which is called epi-convergence. In Euclidean spaces, convergence of a sequence of epigraphs in the Attouch-Wets sense is equivalent to convergence in the sense of Painlevé-Kuratowski (Beer, 1993, Theorem 5.2.10). That is, the convergence of a sequence of sets $\{A_n\}$ to a limit set A is equivalent to the following two conditions: (i) for all $x \in A$ there is a sequence $\{x_n\}$ with $x_n \in A_n$ for each n such that $x_n \to x$ and (ii) for each sequence $\{x_n\}$ with $x_n \in A_n$ for each n, there is a subsequence x_{n_k} such that $x_{n_k} \to x \in A$.

Definition 1 (Epi-convergence)**.** Suppose that $\{f_n\}$ and f are lower semicontinuous functions mapping $\mathbb{R}^p$ to $\mathbb{R}$. Then $\{f_n\}$ *epi-converges* to f, written $f_n \overset{\text{epi}}{\to} f$, if and only if $\lim_{n\to\infty} d_{\mathcal{F}}(\text{epi}f_n, \text{epi}f) = 0$.

Epi-convergence is useful because it is an intermediate concept between the pointwise convergence of functions, which is too weak a concept to describe uniform asymptotic behavior, and uniform convergence, which is too strong because it does not allow for infinite function values, which is the technical means by which constraints are modeled. When the functions involved in epi-convergence are themselves all bounded, epi-convergence is equivalent to uniform convergence on the space of bounded functions equipped with the supremum metric. Bücher et al. (2014) Sections 2, 3 and Appendix B offer a compact introduction to related issues and Royset (forthcoming) additionally provides results that are relevant to statistical properties.

The concepts of convergence in probability and weak convergence in this space have the usual definitions for any (pseudo)metric space (van der Vaart and Wellner, 1996, for example). Consider a sequence of function $\{f_n\}$ and a limiting f. Then f_n epi-converges in probability to f if $\lim_n \mathrm{P}\{d_{\mathcal{F}}(\text{epi}f_n, \text{epi}f) > \epsilon\} \to 0$ for each $\epsilon > 0$, where probability is implicitly outer probability to avoid measurability issues. This will be denoted $f_n \overset{p}{\to} f$. Weak convergence of f_n to f is equivalent to several criteria (van der Vaart and Wellner, 1996, Theorem 1.3.4) interpreted in the pseudometric space $(\mathcal{F}, d_{\mathcal{F}})$. When processes are uniformly bounded weak epi-convergence is equivalent to using the space of bounded functions with the uniform metric, that is, it is equivalent to the weak convergence in $\ell^\infty(\mathbb{R}^p)$ as described in van der Vaart and Wellner (1996). In the applications of this theory below, the convergence of the quantile regression objective function occurs in the general epi-convergent sense, while for other bounded objects like quantile regression solutions and inference processes, convergence is equivalent to convergence in the space of bounded functions, and for this reason, epi-convergence in probability is denoted $f_n \overset{p}{\to} f$ and weak epi-convergence is written $f_n \rightsquigarrow f$.

The following assumptions and definitions on the data generating process are fairly standard in the quantile regression literature (Koenker, 2005; Angrist et al., 2006).

A1 For each $i = 1, \dots n$ the sample observations $\{Y_i, X_i\}_{i=1}^n$ are iid, where $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^p$. Assume X contains an intercept and

$$Q_{Y|X}(\tau|X = X_i) = X_i^\top \beta_0(\tau). \tag{3}$$

A2 $F_{Y|X}$ has uniformly continuous density $f_{Y|X}$ which satisfies $0 < f_{Y|X}(y|X) < \infty$ over the set $\{y : 0 < F_{Y|X}(y|X) < 1\}$ uniformly in X. Let $\mathcal{T} \subseteq [\epsilon, 1-\epsilon]$ for $\epsilon \in (0, 1/2)$, and assume $\mathcal{T} \subset \{y : 0 < F_{Y|X}(Y|X) < 1\}$ uniformly in X.

A3 $D = \mathrm{E}\left[XX^\top\right]$ is a positive definite matrix.

A4 $H(\tau) = \mathrm{E}\left[f_{Y|X}(Q_{Y|X}(\tau|X))XX^\top\right]$ is a positive definite matrix.

A5 $\mathrm{E}\left[\|X\|^4\right] < \infty$.

Assumption **A5** could be weakened to a bound on the $2+\eta$ moment of X for some $\eta > 0$ if one were only concerned about the objective surface and coefficient estimates, but a bound on the fourth moment of X is needed for the uniform consistency of covariance matrix estimators (Angrist et al., 2006, Section A.1.4). The next two assumptions about the constraints imposed on the true coefficient function β_0 are new additions to the list. Note that β is a function of $\tau \in \mathcal{T}$ so the constraint set $\mathcal{B}$ is a set-valued correspondence indexed by $\tau \in \mathcal{T}$. To discuss dependence on a specific $\tau \in \mathcal{T}$ call $\beta(\tau) \in \mathbb{R}^p$ marginal regression coefficients and $\mathcal{B}(\tau)$ marginal constraint sets. Finally, recall that if $C \subseteq \mathbb{R}^k$ is a cone with vertex at v, then $x \in C$ implies $\lambda(x-v)+v \in C$ for all $\lambda \geq 0$.

A6 $\beta_0(\cdot)$ (as a function in $\mathcal{T}$) lies in a closed constraint set $\mathcal{B}(\cdot) \subseteq \mathbb{R}^p \times \mathcal{T}$. For each $\tau \in \mathcal{T}$, the marginal $\mathcal{B}(\tau)$ is convex. As $n \to \infty$, $d_{\mathcal{F}}(\sqrt{n}(\mathcal{B}-\beta_0), T_{\mathcal{B}}(\beta_0)) \to 0$ for some $T_{\mathcal{B}}(\beta_0) \subseteq \mathbb{R}^p \times \mathcal{T}$.

A7 $T_{\mathcal{B}}(\beta_0)$ is a product set: $T_{\mathcal{B}}(\beta_0) = \mathcal{C} \times \mathcal{T}$ where $\mathcal{C} \subseteq \mathbb{R}^p$ is a convex cone with vertex at $\mathbf{0}_p$.

Assumptions **A6** and **A7** require that the recentered and scaled constraint set $\sqrt{n}(\mathcal{B}-\beta_0) \in \mathcal{F}(\mathbb{R}^p \times \mathcal{T})$ can be approximated locally around the true β_0 by a tangent set that is invariant in τ. The lack of dependence on τ may be restrictive, but it avoids situations where the tangent set is not stable over $\mathcal{T}$ that can lead to discontinuities in the distributions of stochastic processes used for testing[1]. Assumption **A7** generalizes the approximation of constraint sets by cones in the sense of Chernoff (Geyer, 1994, Theorem 2.1). Note that $T_{\mathcal{B}}$ is not itself a cone because in one direction it is an interval, but by assumption the constraint set is marginally conical at each $\tau \in \mathcal{T}$. The constraint set is assumed to be fixed but given the general applicability of the epi-convergence concept, assumption **A6** could be relaxed to a sequence of constraints that converge appropriately to $T_{\mathcal{B}}(\beta_0)$ (see for example Geyer (1994) about regularity in the convergence of sets to their tangent cones for a fixed quantile level).

It is convenient to reparameterize $\tilde{\beta}$ as $\tilde{\delta} = \sqrt{n}(\tilde{\beta} - \beta_0(\cdot))$. If $\tilde{\delta}$ is the minimizer of the objective function defined below then $\tilde{\beta} = \tilde{\delta}/\sqrt{n} + \beta_0$ is the minimizer of the usual quantile regression objective function.

Theorem 2.1. *For $(\delta, \tau) \in \mathbb{R}^p \times \mathcal{T}$, define*

$$Z_n(\delta,\tau) = \begin{cases} \sum_{i=1}^n \left(\rho_\tau(U_i - X_i^\top \delta/\sqrt{n}) - \rho_\tau(U_i)\right), & \delta \in \sqrt{n}(\mathcal{B}(\tau) - \beta_0(\tau)) \\ +\infty & \textit{otherwise} \end{cases}$$

[1]Here is a simple example in which τ-dependence of the tangent set is problematic: let $\mathcal{B}(\tau) = \mathbb{R}^2_+$ for all τ and $\beta_0(\tau) = [(1/2-\tau)_+, (\tau-1/2)_+]$. Then the set $T_{\mathcal{B}}(\beta_0) \subset \mathbb{R}^2 \times \mathcal{T}$ (recall $\mathcal{T} = [\epsilon, 1-\epsilon]$) and $d_{\mathcal{F}}(\sqrt{n}(\mathcal{B}-\beta_0), T_{\mathcal{B}}(\beta_0)) \to 0$ with

$$T_{\mathcal{B}}(\beta_0) = (\mathbb{R} \times \mathbb{R}_+ \times [\epsilon, 1/2)) \cup \left(\mathbb{R}^2_+ \times \{1/2\}\right) \cup (\mathbb{R}_+ \times \mathbb{R} \times (1/2, 1-\epsilon]).$$

where the quantile-specific errors $U_i = Y_i - Q_{Y|X}(\tau|X_i)$*, and define*

$$Z(\delta,\tau) = \begin{cases} -\delta^\top G(\tau) + \frac{1}{2}\delta^\top H(\tau)\delta, & \delta \in \mathcal{C} \\ +\infty & \textit{otherwise} \end{cases}$$

where G *is a Gaussian process on* $[0,1]^p$ *with mean zero and* $\mathrm{Cov}(G(s), G(t)) = (s \wedge t - st)D$*. Under Assumptions* ***A1-A7****,* $Z_n \rightsquigarrow Z$ *in* $(\mathcal{F}, d_{\mathcal{F}})$*.*

Theorem 2.1 expresses the quantile regression objective function as a stochastic process in tau, or viewed in another light, as a parametric family of constrained minimization problems in δ with τ as the parameter. The proof of Theorem 2.1 uses results in Kato (2009) about argmin processes and epi-convergence results of Geyer (1994) and Rockafellar and Wets (1997). Theorem 2.1 shows that the objective function has a uniform quadratic approximation that applies to each value of τ and simultaneously provides the main ingredient in the $\sqrt{n}$-consistency result (shown below) that is often stated as a preliminary lemma in the constrained inference literature, for example in Silvapulle and Sen (2005, Lemma 4.2.3).

For each n define the solution mapping $\tilde{\delta}_n(\cdot) : \mathcal{T} \to \mathbb{R}^p$ for each τ by

$$\tilde{\delta}_n(\tau) = \operatorname*{argmin}_{\delta \in \sqrt{n}(\mathcal{B}(\tau) - \beta_0(\tau))} Z_n(\delta, \tau).$$

Analogously define $\tilde{\delta} : \mathcal{T} \to \mathbb{R}^p$ for each τ by

$$\tilde{\delta}(\tau) = \operatorname*{argmin}_{\delta \in \mathcal{C}} Z(\delta, \tau).$$

Theorem 2.2 below shows that the sequence of finite-sample minimizer processes $\tilde{\delta}_n$ converges weakly to the asymptotic minimizer $\tilde{\delta}$, and that the value function processes converge weakly as well. Because Z depends continuously on τ, the sequence of minimizers can be interpreted as a sequence of stochastic processes in $\ell^\infty(\mathbb{R}^p)$ that converges weakly to a continuous limit (Rockafellar and Wets, 1997, Theorem 2.6 and Corollary 7.43).

Theorem 2.2. *Under Assumptions* ***A1-A7***

(a) $\tilde{\delta}_n \rightsquigarrow \tilde{\delta}$*.*

(b) $Z_n(\tilde{\delta}_n, \cdot) \rightsquigarrow Z(\tilde{\delta}, \cdot)$*.*

By reparameterizing from $\tilde{\delta}_n$ back to $\tilde{\beta}$, part (a) implies that $\tilde{\beta}$ is a uniformly $\sqrt{n}$-consistent estimator of β_0. This argument follows the same order as in Kato (2009) and is slightly different from other analyses that start by showing the consistency of the estimator and then show its asymptotic normality. Note also that the constrained estimator can not in general have an asymptotically normal distribution, because of the potential of binding constraints.

Theorems 2.1 and 2.2 apply to a stochastic process of minimization problems. A little manipulation allows one to characterize the distribution of the limiting $\tilde{\delta}$ in a more enlightening way. Lemma 2.3 focuses on characterizing the objective function and the value function at a single quantile, making it similar to many well known results in the constrained inference literature. The main difference is that the characterizations below hold uniformly in τ.

Lemma 2.3. *Assume **A1-A7**. Define the stochastic process $W \in C(\mathcal{T})^p$ by $W(\cdot) = H^{-1}(\cdot)G(\cdot)$, where G was defined in Theorem 2.1. Then*

(a) $Z(\delta, \tau) = \frac{1}{2}\left((\delta - W(\tau))^\top H(\tau)(\delta - W(\tau)) - G^\top(\tau)H^{-1}(\tau)G(\tau)\right)$

(b) $\tilde{\delta}(\tau) = \text{argmin}_{\delta \in \mathcal{C}} (\delta - W(\tau))^\top H(\tau)(\delta - W(\tau))$

(c) $Z(\tilde{\delta}(\tau), \tau) = -\frac{1}{2}\tilde{\delta}^\top(\tau)H(\tau)\tilde{\delta}(\tau)$.

If $\beta_0(\tau)$ is in the interior of $\mathcal{B}(\tau)$, then $\mathcal{C} = \mathbb{R}^p$ and the optimum of the resulting unconstrained quadratic program can be found analytically, and $\delta(\tau) = W(\tau) \sim \mathcal{N}\left(\mathbf{0}, \tau(1-\tau)H^{-1}(\tau)DH^{-1}(\tau)\right)$, the asymptotic distribution for the unconstrained quantile regression estimator.

3 Constrained regression rankscore processes

The quantile regression estimation problem is conveniently posed as an L_1 minimization problem and as such it has a dual maximization problem. Gutenbrunner and Jurečková (1992) showed that the solution to this dual problem, taken as a stochastic process in τ, is useful for inference and L-estimation, and they called it the *regression rankscore* process. Roughly speaking, the vector of τ^{th} regression rankscores can be fashioned into a score statistic for the τ^{th} quantile regression coefficient estimate. Gutenbrunner et al. (1993) extended this methodology to tests for linear restrictions in quantile regression models and Koenker and Machado (1999) extended this to rankscore processes for uniform inference over an interval of quantile levels. Inference using constrained regression rankscores — that is, the dual solution to the constrained quantile regression problem — have not been previously considered. Here the constraint set is specialized to a set of linear inequality constraints and properties of the dual solution that are useful for inference on constrained quantile regression coefficients are explored.

Given a sample of size n let $Y \in \mathbb{R}^n$ be a vector of response observations and $X \in \mathbb{R}^{n \times p}$ collect the observations with each $X_i^\top$ laying in the corresponding row of X. Specialize the constraint set $\mathcal{B}(\tau)$ to the marginally linear form $\mathcal{B}(\tau) = \{\beta \in \mathbb{R}^p : R\beta \geq r\}$, where $R \in \mathbb{R}^{q \times p}$ has rank $q \leq p$ and $r \in \mathbb{R}^q$. Estimation of quantile regression coefficients under linear inequality constraints has been considered in Portnoy and Koenker (1997); Koenker and Ng (2005)[2]. Specifically, the (primal) constrained quantile regression problem for any quantile level τ can be written as

$$\min_{u,v,b} \left\{\tau \mathbf{1}_n^\top u + (1-\tau)\mathbf{1}_n^\top v : Y = Xb + u - v, Rb \geq r, u, v \geq \mathbf{0}_n\right\}$$

[2]The primal solution can be found using the R package `quantreg` — specifically, the estimator is implemented in `quantreg`'s functions `rq.fig.fnc` and `rq.fit.sfnc`.

letting u and v represent positive and negative quantile residuals in estimation. Define

$$\tilde{X} = \begin{bmatrix} X \\ R \end{bmatrix} \in \mathbb{R}^{(n+q)\times p}, \quad \tilde{Y} = \begin{bmatrix} Y \\ r \end{bmatrix} \in \mathbb{R}^{n+q}.$$

Then the dual of the constrained minimization problem can be written (see the Appendix for more details)

$$\max_{\lambda} \left\{ \tilde{Y}^\top \lambda : \tilde{X}^\top \lambda = \mathbf{0}_{p+q}, \lambda \in [\tau - 1, \tau]^n \times \mathbb{R}_+^q \right\}. \tag{4}$$

The solution to this problem (as a function of τ) is called the *constrained regression rankscore* process $\tilde{\lambda}_n(\tau) = [\tilde{\lambda}_{1n}^\top(\tau), \tilde{\lambda}_{2n}^\top(\tau)]^\top$. $\tilde{\lambda}_n$ is an $(n+q)$-dimensional stochastic process that can be subdivided into $\tilde{\lambda}_{1n} \in [\tau - 1, \tau]^n$ and $\tilde{\lambda}_{2n} \in \mathbb{R}_+^q$ for each τ. The $\tilde{\lambda}_{1n}$ coordinates act like unconstrained regression rankscores, while the $\tilde{\lambda}_{2n}$ coordinates correspond to Lagrange multipliers that are associated with the inequality constraints imposed in the constrained problem.

Since the quantile regression optimization problem is a linear programming problem with linear constraints, it has a solution at one vertex or a convex combination of several vertices, denoted the *p basic* sample observations or constraints for that solution. To discuss the set of basic elements of $\tilde{X}$ and $\tilde{Y}$, let $\mathcal{H}$ denote the collection of p-element subsets of $\{1, 2, \ldots, n, n+1, \ldots n+q\}$. Each h has a complement in $\mathcal{H}$ denoted $\bar{h} := \{1, \ldots, n+q\}/h$. Let $\tilde{X}(h) := \{\tilde{X}_{i,\cdot} : i \in h\}$ and $\tilde{Y}(h) := \{\tilde{Y}_i : i \in h\}$. Also define $h_1 = h \cap \{1, 2, \ldots n\}$, $\bar{h}_1 = \bar{h} \cap \{1, 2, \ldots n\}$, $h_2 = h \cap \{n+1, \ldots, n+q\}$ and $\bar{h}_2 = \bar{h} \cap \{n+1, \ldots, n+q\}$. h indexes all the basic elements of $\tilde{X}$ and $\tilde{Y}$, while h_1 and h_2 index which elements of the design matrix or constraint matrix are basic.

Because the quantile regression problem is feasible, it has at least one solution (there may in general be a hypersurface of solutions), resulting in the following fact, which mirrors Theorem 3.1 of Koenker and Bassett (1978). Assuming $\mathrm{rank}(\tilde{X}) = p$, the set of solutions to the primal problem (3) has at least one element of the form

$$\tilde{\beta}(\tau) = \tilde{X}(h)^{-1} \tilde{Y}(h)$$

for some $h \in \mathcal{H}$ with $\mathrm{rank}(\tilde{X}(h)) = p$ (the full set of solutions is the convex hull of all such vertex solutions). An important consequence of this solution is that exactly p elements of $(\tilde{X}, \tilde{Y})$ go into the solution $\tilde{\beta}$, not elements in (X, Y). In other words, only observations indexed by h_1 will be fit exactly by the estimated regression plane — $i \in h_1$ implies $Y_i = X_i^\top \tilde{\beta}(\tau)$. Similarly, constraints indexed by h_2 hold exactly — that is, $i \in h_2$ implies $R_{i,\cdot}\tilde{\beta}(\tau) = r_i$. The configuration of any given sample makes the number of binding constraints on $\tilde{\beta}$ random, and because $\mathrm{card}(h_1) + \mathrm{card}(h_2) = \mathrm{card}(h) = p$, the number of exactly-interpolated sample observations is also random, in contrast to the unconstrained quantile regression estimator.

Regression rankscores can be manipulated to provide a variety of different tests (Gutenbrunner et al., 1993, Section 2) but for the inference processes for quantile regression considered below it is convenient to use $\tilde{\lambda}$ directly without reparameterization. The random variable used for inference in the subsequent section uses only the contributions from the $\tilde{\lambda}_{1n}$ subvector. This relies on the fact that $\tilde{\lambda}_{1ni}(\tau)$ is with a high probability identical to $\tau - I(Y_i < X_i^\top \beta_0(\tau))$, which is a subgradient function for the quantile regression objective function marginally at τ.

Lemma 3.1. *Let $\psi_\tau(u) = \tau - I(u < 0)$. Under assumptions **A1**-**A7**,*

$$\sup_{\tau\in\mathcal{T}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \tilde{\lambda}_{1ni}(\tau) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \psi_\tau(Y_i - X_i^\top \tilde{\beta}(\tau)) \right\| = o_P(1).$$

Lemma 3.2 formulates a relationship between the subgradient of the objective function (at $\tilde{\beta}$ and β_0) and the coefficient estimate. It is used (together with Lemma 3.1) in the next section to provide an asymptotic description for regression rankscore statistics, which are the quantile regression analog of a score statistic.

Lemma 3.2. *Let $\psi_\tau(u) = \tau - I(u < 0)$. Under assumptions **A1**-**A7**,*

$$\sup_{\tau\in\mathcal{T}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \psi_\tau(Y_i - X_i^\top \tilde{\beta}(\tau)) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \psi_\tau(Y_i - X_i^\top \beta_0(\tau)) + H(\tau)\sqrt{n}(\tilde{\beta}(\tau) - \beta_0(\tau)) \right\| = o_P(1).$$

Lemma 3.2 is similar to Theorem 1 of Gutenbrunner and Jurečková (1992) (see also Theorems 3.1-3.3 of Gutenbrunner et al. (1993)) but tailored to the constrained quantile regression estimator. It is also akin to the Bahadur representation — that is, an asymptotically linear representation — for unconstrained quantile regression estimates (Koenker, 2005, Section 4.3), although one should not expect an asymptotic linear form to hold for estimates that do not have an asymptotically normally distribution. The typical Bahadur representation would not have the first term in the previous display, which is negligible when estimates are unconstrained. However, equation (4) implies $X^\top \tilde{\lambda}_{1n}(\tau) = -R^\top \tilde{\lambda}_{2n}(\tau)$. That is, unlike unrestricted regression rankscores, $X^\top \tilde{\lambda}_{1n}$ is not generally equal to zero.

4 Inference processes and their asymptotic distributions

The inference processes considered here were investigated in Koenker and Machado (1999), except that here inequality-constrained estimates are used. As in Section 3, the hypotheses below are simplified to polyhedra defined by linear inequalities. Focusing on just these polyhedral constraints implies some simplification. Specifically let $R \in \mathbb{R}^{q\times p}$ for $q \leq p$ and let $r \in \mathbb{R}^q$. Assume rank$(R) = q$. *Type A* hypotheses have maintained hypothesis $K^A : R\beta_0(\tau) \geq r$ for all $\tau \in \mathcal{T}$ and have null and alternative hypotheses

$$\begin{aligned} H_0^A &: R\beta_0(\tau) = r \quad \text{for all } \tau \in \mathcal{T} \\ H_1^A &: R\beta_0(\tau_0) > r \quad \text{for some } \tau_0 \in \mathcal{T}. \end{aligned}$$

Type B hypotheses have maintained hypothesis $K^B : \beta(\tau) \in \mathbb{R}^p$ for all $\tau \in \mathcal{T}$ and null and alternative

$$\begin{aligned} H_0^B &: R\beta_0(\tau) \geq r \quad \text{for all } \tau \in \mathcal{T} \\ H_1^B &: R\beta_0(\tau_0) \ngeq r \quad \text{for some } \tau_0 \in \mathcal{T}. \end{aligned}$$

Common estimators are used to construct a variety of inference processes below, under either set of hypotheses. There are three relevant coefficient estimators, labeled

$$\bar{\beta}(\tau) = \operatorname*{argmin}_{b:Rb=r} \sum_i \rho_\tau(Y_i - X_i^\top b)$$
$$\tilde{\beta}(\tau) = \operatorname*{argmin}_{b:Rb\geq r} \sum_i \rho_\tau(Y_i - X_i^\top b)$$
$$\hat{\beta}(\tau) = \operatorname*{argmin}_{b\in\mathbb{R}^p} \sum_i \rho_\tau(Y_i - X_i^\top b).$$

The matrix D can be estimated the same way under Type A or B hypotheses: let

$$D_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top.$$

In order to impose the null hypothesis on test statistics for proper size and power in testing procedures, two different estimates of and H are needed, which only depend on which coefficient estimator is used in the definition. Define

$$\bar{H}_n(\tau) = \frac{1}{nh_n}\sum_{i=1}^n K((Y_i - X_i^\top \bar{\beta}(\tau))/h_n) X_i X_i^\top,$$
$$\tilde{H}_n(\tau) = \frac{1}{nh_n}\sum_{i=1}^n K((Y_i - X_i^\top \tilde{\beta}(\tau))/h_n) X_i X_i^\top,$$

where K is a kernel function and $\{h_n\}$ is a bandwidth sequence such that $h_n \to 0$ and $\sqrt{n}h_n \to \infty$. H_n using the unrestricted $\hat{\beta}$ was first suggested in Powell (1986), and given assumptions **A3**-**A5**, section A.1.4 of Angrist et al. (2006) shows that H_n is a uniformly consistent estimator of H. Finally, the estimated variance function of the Gaussian process W defined in Lemma 2.3 will be used as a weighting matrix in several places. This estimate also depends on the maintained hypothesis via coefficient estimate. Let

$$\tilde{\Sigma}_n(\tau) = \tau(1-\tau)\tilde{H}_n^{-1}(\tau) D_n \tilde{H}_n^{-1}(\tau),$$
$$\bar{\Sigma}_n(\tau) = \tau(1-\tau)\bar{H}_n^{-1}(\tau) D_n \bar{H}_n^{-1}(\tau).$$

Denote the limit in probability of these estimators as Σ. When the data generating process is homoskedastic, this assumption and the linear model assumption imply that the distribution of Y follows a location-scale model: there is some density function f_0 such that $f_{Y|X}(y|x) = \sigma^{-1} f_0((y - x^\top\beta)/\sigma)$, and some quantile function Q_0 such that $Q_{Y|X}(\tau|x) = x^\top\beta + \sigma Q_0(\tau)$. Then Σ is simplified by the fact that $H(\tau) = \sigma^{-1} f_0(Q_0(\tau))D$. This means that the matrix Σ_n can be reduced to a special case: let

$$\Sigma_{n0}(\tau) = \tau(1-\tau)\sigma^2 \widehat{f_0^{-2}(Q_0(\tau))} D_n^{-1}.$$

Under the assumption of homoskedasticity one may estimate the inverted density (or sparsity) function in Σ_{n0} using the proposal of Hendricks and Koenker (1992) (see Koenker and Machado (1999) for more details). Analogous to the more general case discussed above, denote the limit in probability of this estimator as Σ_0.

Before discussing inference processes it is necessary to define their limiting process. To this end, it may help to recall some basic facts about order-restricted inference for vector-valued parameters. In the classical literature on order-restricted inference the $\bar{\chi}^2$ statistic is a common limiting distribution for likelihood ratio statistics with one estimate subject to inequality constraints. Suppose that under the maintained hypothesis $K : R\theta \geq r$ for some $\theta \in \mathbb{R}^p$, $R \in \mathbb{R}^{q\times p}$ with $q \leq p$ and $r \in \mathbb{R}^q$, we wish to test $H_0 : R\theta = r$ against $H_1 : R\theta > r$. Suppose furthermore that given an observed sample, k of q constraints actually hold with equality. Then asymptotic theory implies that under the null hypothesis, the likelihood ratio statistic and related statistics converge to a χ^2_k random variable. However, when estimates are made subject to inequality constraints, the sample realizations randomly determine which constraints bind. This implies that the asymptotic distribution under the null hypothesis is a mixture of χ^2_k variables for k ranging from 0 to q, the total number of constraints. This limiting distribution is commonly referred to as a $\bar{\chi}^2$ distribution (Silvapulle and Sen, 2005, p. 75). The weights in the mixture, $\{w_k\}_{k=1}^q$, can be derived analytically when the constraints are simple, but simulation is the only general way to estimate tail probabilities from a $\bar{\chi}^2$ distribution (Silvapulle and Sen, 2005, Sections 3.3-3.6).

In order to uniformly describe the behavior of the quantile regression process under inequality constraints, a stochastic process that is analogous to the $\bar{\chi}^2$ random variable is required. Let B_k be a continuous Gaussian process in $\mathbb{R}^k$ with mean zero and covariance $\mathrm{Cov}(B_k(s), B_k(t)) = (s \wedge t - st) \times I_k$ for $s, t \in [0, 1]$, that is, a vector-valued process with independent Brownian bridge processes at each coordinate. Then define a scaled Bessel bridge process of order k as $Q_k(\tau) = \|B_k(\tau)\|/\sqrt{\tau(1-\tau)}$. Finally, define Q_0 as the function that is identically zero over the unit interval. For any integer $k > 0$ and fixed τ, $Q^2_k(\tau) \sim \chi^2_k$. Q^2_k processes with a fixed k were used in Koenker and Machado (1999) for inference processes in unconstrained quantile regression. The asymptotic distribution of inequality-constrained quantile regression processes is a mixture of Q^2_k processes for $k = 0, \ldots q$.

Definition 2 ($\bar{Q}^2$ process)**.** Let $\bar{Q}^2 : [0, 1] \to \mathbb{R}$ be a stochastic process

$$\bar{Q}^2(\tau) = \sum_{k=0}^{q} w_k(\tau) Q^2_k(\tau),$$

where Q_0 is a function identically equal to zero, for $k = 1, 2, \ldots q$, Q_k is a scaled Bessel bridge of order k and $w : [0, 1] \to [0, 1]^q$ is a weight function. Each coordinate of w satisfies $w_k(\tau) \geq 0$ and $\sum_{i=1}^q w_k(\tau) = 1$ for all τ.

A $\bar{Q}^2$ process has marginal $\bar{\chi}^2$ distributions. The weight function depends on the constraints and the covariance function of the process and it can generally be found by simulation.

Given the two different sets of hypotheses, two particular $\bar{Q}^2$ processes are relevant for inference. Specifying them is equivalent to describing their weighting functions, which are functions in $\mathcal{T}$. For type

A hypotheses, let $\mathcal{M} = \{\delta \in \mathbb{R}^p : R\delta = \mathbf{0}_q\}$. Then define the process $\bar{Q}^2_A$ by its marginal distributions

$$\mathrm{P}\left\{\bar{Q}^2_A(\tau) \le c\right\} = \sum_{k=0}^{q} w_k(R\Sigma(\tau)R^\top, \mathcal{M})\mathrm{P}\left\{\chi^2_k \le c\right\}, \quad c > 0,$$

where the k^{th} coordinate of the weight function, $w_k(V, S)$ is the probability that k constraints bind in the minimization problem $\min_{X \in S} X^\top V^{-1} X$ where $X \sim \mathcal{N}(\mathbf{0}_q, V)$ and S is a region defined by a polyhedral constraint cone. This means that the distribution depends on the constraints and covariance function of the coefficient process; it is not a pivotal distribution. Type B hypotheses require a slightly different limiting distribution: first let $\mathcal{C} = \{\delta \in \mathbb{R}^p : R\delta \ge \mathbf{0}_q\}$, and define the $\bar{Q}^2_B$ process with marginal distributions

$$\mathrm{P}\left\{\bar{Q}^2_B(\tau) \le c\right\} = \sum_{k=0}^{q} w_{q-k}(R\Sigma(\tau)R^\top, \mathcal{C})\mathrm{P}\left\{\chi^2_k \le c\right\}, \quad c > 0.$$

In the definition of this process the weight coordinates w_k are defined as in the type-A case. As will be seen below, the $\bar{Q}^2_B$ process is only useful under the sub-hypothesis of the type-B null where the constraints hold with equality.

An analog of the likelihood ratio test generalized to M-estimators was suggested by Ronchetti (1982) and denoted a ρ test. A quantile regression ρ test (for a single τ) is a true likelihood ratio test statistic only under the assumption that the error distribution follows a special asymmetric Laplace density; generally it simply measures the drop in the value function associated with more- and less-constrained quantile regression estimation problems. Two closely related ρ processes for inference over $\mathcal{T}$ were defined in Koenker and Machado (1999), for when the scale of the error distribution is respectively known or unknown. Let $\bar{V}_n(\tau) = \sum_i \rho_\tau(Y_i - X_i^\top \bar{\beta})$ with analogous definitions for $\tilde{V}_n(\tau)$ and $\hat{V}(\tau)$. Unfortunately, these ρ processes can only be defined for homoskedastic data, as specified in assumption **B1** below. The following regularity conditions are needed to assure that the ρ processes have well defined asymptotic properties.

B1 The conditional quantiles of Y given X are described by $Q_{Y|X}(\tau|X) = X^\top\beta + F^{-1}(\tau)$ for each $\tau \in \mathcal{T}$.

B2 $\tilde{\sigma}(\tau) := n^{-1}\tilde{V}_n(\tau)$ or $\hat{\sigma}(\tau) := n^{-1}V_n(\tau)$ are uniformly consistent estimators of $\mathrm{E}[\rho_\tau(\tau)] := \sigma(\tau)$ under maintained hypotheses K^A and K^B respectively.

Assumption **B1** maintains that the data is generated according to a homoskedastic linear model $Y = X\beta + U, U \sim F$, where β is a τ-invariant vector. This implies the model of the conditional quantiles of Y given X is the location-shift model implicit in simple mean regression models. This is quite restrictive, but the other inference processes defined below do not require this assumption and have the same asymptotic behavior, as described in subsequent theorems.

For the known-scale (set to 1 without loss of generality) case, define

$$L_n^A(\tau) = \frac{2\widehat{f_0(Q_0(\tau))}}{\tau(1-\tau)}\left(\bar{V}_n(\tau) - \tilde{V}_n(\tau)\right)$$
$$L_n^B(\tau) = \frac{2\widehat{f_0(Q_0(\tau))}}{\tau(1-\tau)}\left(\tilde{V}_n(\tau) - \hat{V}_n(\tau)\right).$$

When the scale of the error distribution is unknown, use

$$\Lambda_n^A(\tau) = \frac{2n\tilde{\sigma}(\tau)\widehat{f_0(Q_0(\tau))}}{\tau(1-\tau)}\log\left(\bar{V}_n(\tau)/\tilde{V}_n(\tau)\right)$$
$$\Lambda_n^B(\tau) = \frac{2n\hat{\sigma}(\tau)\widehat{f_0(Q_0(\tau))}}{\tau(1-\tau)}\log\left(\tilde{V}_n(\tau)/\hat{V}_n(\tau)\right).$$

In the following theorems, all inference processes are bounded functions, and so weak convergence (denoted $\rightsquigarrow$) is the standard weak convergence concept for bounded functions used in van der Vaart and Wellner (1996).

Theorem 4.1. *Make Assumptions **A1**-**A7**, **B1**-**B2**. Then*

(a) Under H_0^A, $\Lambda_n^A(\tau) = L_n^A(\tau) + o_P(1)$ and $\Lambda_n^B(\tau) = L_n^B(\tau) + o_P(1)$, uniformly over $\mathcal{T}$.

(b) Under H_0^A, $L_n^A \rightsquigarrow \bar{Q}_A^2$, where $\Sigma = \Sigma_0$.

(c) Let $\sqrt{n}(\bar{\beta} - \beta_0) \rightsquigarrow \bar{\delta}$ and $\sqrt{n}(\tilde{\beta} - \beta_0) \rightsquigarrow \tilde{\delta}$. Then $\tilde{\delta}^\top \Sigma_0^{-1}\tilde{\delta} - \bar{\delta}^\top \Sigma_0^{-1}\bar{\delta} \sim \bar{Q}_A^2$.

(d) Let $\sqrt{n}(\bar{\beta} - \beta_0) \rightsquigarrow \bar{\delta}$ and $\sqrt{n}(\tilde{\beta} - \beta_0) \rightsquigarrow \tilde{\delta}$. Suppose that δ can logically be divided coordinate-wise in two sub-processes $\delta = [\delta_1^\top, \delta_2^\top]^\top$ with $\delta_1 : \mathcal{T} \to \mathbb{R}^{p-q}$ and $\delta_2 : \mathcal{T} \to \mathbb{R}^q$, and let $R = [\mathbf{0}_{q\times(p-q)}, R_2]$ where $R_2 \in \mathbb{R}^{q\times q}$ is a rank-q matrix, so that δ_1 is the limit process for the p−q unrestricted coordinates under H_0^A. Then $\bar{\delta}_2^\top \left(R\Sigma_0 R^\top\right)^{-1}\bar{\delta}_2 - \tilde{\delta}_2^\top \left(R\Sigma_0 R^\top\right)^{-1}\tilde{\delta}_2 \sim \bar{Q}_A^2$.

(e) Under H_0^B, $L_n^B \rightsquigarrow \bar{Q}^2$ with some weight function w. Under the hypothesis H_0^A, $L_n^B \rightsquigarrow \bar{Q}_B^2$, where $\Sigma = \Sigma_0$.

(f) Let $\sqrt{n}(\tilde{\beta} - \beta_0) \rightsquigarrow \tilde{\delta}$. Under H_0^A, $W^\top \Sigma_0^{-1} W - \tilde{\delta}^\top \Sigma_0^{-1}\tilde{\delta} \sim \bar{Q}_B^2$.

In Theorem 4.1, H_0^A must apply to both type A processes (which is usual, in part b) and type B processes (in part e) to obtain a fully-specified null distribution. The type B problem has a composite null hypothesis, and so only under the restricted subhypothesis H_0^A within the larger type B null parameter space does the ρ process have an asymptotic $\bar{Q}_B^2$ characterization. Using the $\bar{Q}_B^2$ distribution for asymptotic inference when H_0^A is not true can distort rejection probabilities under the null and lower the power of uniform inference procedures. Raising the size and power of such tests is the focus of a body of ongoing research represented for example by Linton et al. (2010), Lee et al. (2013), Chernozhukov et al. (2013), Andrews and Shi (2017), Chernozhukov et al. (forthcoming). Roughly speaking, in this literature one estimates which constraints bind and conducts inference over this set. This sort of inference is very different from the asymptotics discussed here and is beyond the scope of this paper.

Also, the restriction of a set of linear hypotheses and a linear model for conditional quantiles may be seen as rather restrictive, and could potentially be generalized, although when both the model and hypotheses are allowed to be nonlinear (for example, $\mathcal{B}$ could be some sort of differentiable surface), great care must be taken to ensure that tests are conducted appropriately — the pitfalls of conducting inference under nonlinear restrictions using nonlinear models are illustrated succinctly in Wolak (1991) for the vector (that is, non-process) case.

Parts (c) and (d) of Theorem 4.1 may be useful for simulating the asymptotic distribution of the process for type A processes. Particularly, the form of the null in part (d) relies on inequality restricted parameter values that may take less time to find via quadratic programming problem than the full parameter vector.

The major drawback of the ρ processes above is that they require homoskedasticity and therefore have limited appeal in the quantile regression context. Wald processes can be defined when the data are heteroskedastic, allowing for a richer model of the conditional quantiles of the response. A Wald process is a measure of the distance between the hypothesized β and an estimate under the maintained hypothesis as a process over τ. Specifically, for each $\tau \in \mathcal{T}$ define

$$W_n^A(\tau) = n(\bar{\beta}(\tau) - \tilde{\beta}(\tau))^\top R^\top \left(R\bar{\Sigma}_n(\tau)R^\top\right)^{-1} R(\bar{\beta}(\tau) - \tilde{\beta}(\tau))$$
$$W_n^B(\tau) = n(\tilde{\beta}(\tau) - \hat{\beta}(\tau))^\top R^\top \left(R\tilde{\Sigma}_n(\tau)R^\top\right)^{-1} R(\tilde{\beta}(\tau) - \hat{\beta}(\tau)).$$

The following theorem describes the asymptotic properties of these statistics.

Theorem 4.2. *Make assumptions **A1**-**A7**. Then*

(a) Under H_0^A, $W_n^A \rightsquigarrow \bar{Q}_A^2$.

(b) Under H_0^B, $W_n^B \rightsquigarrow \bar{Q}^2$ with some weight function w. Under H_0^A, $W_n^B \rightsquigarrow \bar{Q}_B^2$.

Note that inference using the Wald process requires no assumption about homoskedasticity as the ρ-processes do, and the limiting distribution is the same as for those processes (except with the more general Σ replacing Σ_0). Because the limit process is otherwise identical, parts (c), (d) and (f) of Theorem 4.1 also apply to the asymptotic limit of Wald processes.

Rankscore processes need the same assumptions as those made for Wald inference. For a type A hypothesis, the restricted model involves an equality constraint and can therefore be estimated via reparameterization using the standard quantile regression estimator (Gutenbrunner et al., 1993, remark on p. 312). Let

$$\bar{S}_n(\tau) = \bar{H}_n^{-1}(\tau) X^\top \bar{\lambda}_n(\tau)/\sqrt{n}$$
$$\tilde{S}_n(\tau) = \tilde{H}_n^{-1}(\tau) X^\top \bar{\lambda}_n(\tau)/\sqrt{n},$$

where $\bar{\lambda}_n(\tau)$ is the regression rankscore process under the assumption that the constraints bind. Next define a quadratic form based on normalized differences between less- and more restricted regression rankscore processes:

$$\begin{aligned}
T_n^A(\tau) &= \left(\bar{S}_n(\tau) - \tilde{S}_n(\tau)\right)^\top R^\top \left(R\bar{\Sigma}_n(\tau)R^\top\right)^{-1} R\left(\bar{S}_n(\tau) - \tilde{S}_n(\tau)\right) \\
T_n^B(\tau) &= \tilde{S}_n^\top(\tau)R^\top \left(R\tilde{\Sigma}_n(\tau)R^\top\right)^{-1} R\tilde{S}_n(\tau).
\end{aligned}$$

Rankscore statistics in constrained problems may require two estimates, instead of just one estimate under the null, because of finite-sample conditions that may cause some constraints to be binding while others are slack, as was alluded to in the discussion around Lemma 3.2. Under the type A maintained hypothesis, the score may not necessarily be exactly equal to zero and this leads to the use of $\tilde{S}_n$ in the formula for T_n^A. Note that T_n^B implicitly subtracts the unrestricted $\hat{S}_n$ from $\tilde{S}_n$ because $\hat{S}_n(\cdot) \equiv \mathbf{0}_p$.

Theorem 4.3. *Under assumptions **A1-A7**, $T_n^A(\tau) = W_n^A(\tau) + o_P(1)$ and $T_n^B(\tau) = W_n^B(\tau) + o_P(1)$ uniformly in τ.*

Therefore the regression rankscore processes result in the same asymptotic inferences as the Wald processes under the same assumptions. Under the restriction of homoskedasticity, density-related terms in H and Σ cancel and it can be seen that the regression rankscore test statistic does not depend on the response distribution. However, under heteroskedasticity it will still be necessary to estimate H. Finite-sample behavior of these processes could, of course, be different, as well as finite-sample behavior of test statistics derived from these processes using supremum or L^2 norms, for example.

5 Conclusion

I describe asymptotic distributions for constrained quantile regression objective processes and constrained quantile regression coefficient estimate processes. Asymptotics for inference processes similar to Koenker and Machado (1999) can be derived from the properties of the asymptotic objective surface, coefficient estimates and the constrained regression rankscore process. For a single quantile level the test statistics here are related to well-known results in the constrained inference literature. However, uniform hypotheses for constrained quantile regression coefficients over a set of quantile levels are also derived. There opportunities to generalize these results, in particular using the recent literature on inference for many moment inequalities, and also these methods could be extended to conduct inference in more complex models of conditional quantiles.

A Proofs

Proof of Theorem 2.1. Let $U_i := Y_i - Q_{Y|X}(\tau|X_i) = Y_i - X_i^\top \beta_0(\tau)$ and for a fixed τ let $\delta = \sqrt{n}(\beta - \beta_0(\tau))$ for $\beta \in \mathbb{R}^p$. Using Knight's identity (Knight, 1998) write the finite part of the objective function process for $(\delta, \tau) \in \mathbb{R}^p \times \mathcal{T}$ as

$$\sum_i \left(\rho_\tau(U_i - X_i^\top \delta/\sqrt{n}) - \rho_\tau(U_i)\right) = \delta^\top Z_{1n}(\tau) + Z_{2n}(\delta, \tau)$$

where

$$Z_{1n}(\tau) = \frac{-1}{\sqrt{n}} \sum_i X_i \left(\tau - I(U_i < 0)\right)$$

and

$$Z_{2n}(\delta,\tau) = \sum_i \int_0^{X_i^\top \delta/\sqrt{n}} \sqrt{n} \left(I(U_i \le s) - I(U_i \le 0)\right) \mathrm{d}s. \tag{5}$$

Next define the "intermediate" lower semicontinuous process

$$Y_n(\delta,\tau) = \begin{cases} \delta^\top Z_{1n}(\tau) + \frac{1}{2}\delta^\top H(\tau)\delta & \delta \in \sqrt{n}(\mathcal{B}(\tau) - \beta_0(\tau)) \\ \infty & \text{otherwise.} \end{cases}$$

Consider the difference between Z_n and Y_n. Lemma A.1 implies that for each $\delta \in \sqrt{n}(\mathcal{B} - \beta_0)$,

$$\sup_{\tau\in\mathcal{T}} |Z_n(\delta,\tau) - Y_n(\delta,\tau)| = \sup_{\tau\in\mathcal{T}} |Z_{2n}(\delta,\tau) - \delta^\top H(\tau)\delta/2| = o_P(1),$$

in other words, $Z_n(\delta,\tau) \xrightarrow{p} Y_n(\delta,\tau)$ uniformly in τ as $n \to \infty$ for each $\delta \in \sqrt{n}(\mathcal{B} - \beta)$. Then Lemma 1 of Kato (2009) shows that $\sup_{\tau\in\mathcal{T}} \sup_{\delta\in K} |Z_n(\delta,\tau) - Y_n(\delta,\tau)| \xrightarrow{p} 0$ for all compact $K \in \mathbb{R}^p$.

A useful distance estimate is defined in Theorem 4.36 of Rockafellar and Wets (1997), which shows that a sequence $\{f_n\}$ epi-converges to f if and only if it converges in terms of $\hat{d}_\gamma$, which is not itself a metric or pseudometric, and is defined by

$$\hat{d}_\gamma(\mathrm{epi}f, \mathrm{epi}g) = \min\{\eta > 0 : \mathrm{epi}f \cap \gamma B_{k+1} \subseteq \mathrm{epi}g + \eta B_{k+1} \text{ and } \mathrm{epi}g \cap \gamma B_{k+1} \subseteq \mathrm{epi}f + \eta B_{k+1}\}. \tag{6}$$

Fix a $\gamma \ge 0$ and define the compact set $K_\gamma = \gamma B_{p+2} \cap \sqrt{n}(\mathcal{B} - \beta) \cap \mathcal{T}$. The condition

$$\|Z_n - Y_n\|_{K_\gamma} := \sup_{\delta\in K_\gamma} |Z_n(\delta,\tau) - Y_n(\delta,\tau)| = \eta$$

implies that $Z_n(\delta,\tau) \ge Y_n(\delta,\tau) - \eta$ for all $(\delta,\tau) \in \gamma B_{p+1} \cap \mathcal{T}$ so

$$\mathrm{epi}Z_n \cap \gamma B_{p+2} \subseteq \mathrm{epi}Y_n - \eta \subseteq \mathrm{epi}Y_n + \eta B_{p+2}$$

and similarly

$$\mathrm{epi}Y_n \cap \gamma B_{p+2} \subseteq \mathrm{epi}Z_n - \eta \subseteq \mathrm{epi}Z_n + \eta B_{p+2}.$$

These two conditions imply that

$$\hat{d}_\gamma(\mathrm{epi}Z_n, \mathrm{epi}Y_n) \le \|Z_n - Y_n\|_{K_\gamma}.$$

Then the uniform convergence of Z_n and Y_n on compacta implies

$$0 \le \lim_{n\to\infty} \mathrm{P}\left\{\hat{d}_\gamma(\mathrm{epi}Z_n, \mathrm{epi}Y_n) > \epsilon\right\} \le \lim_{n\to\infty} \mathrm{P}\left\{\|Z_n - Y_n\|_{K_\gamma} > \epsilon\right\} = 0.$$

Theorem 4.36 of Rockafellar and Wets (1997) then implies that $Z_n \xrightarrow{\text{epi}} Y_n$ in probability.

Now define the sequence of "extended indicator" functions $\{w_n\}$, where for each n,

$$w_n(\delta, \tau) = \begin{cases} 0 & \delta \in \sqrt{n}(\mathcal{B} - \beta(\tau)) \\ +\infty & \text{otherwise} \end{cases}$$

and

$$w(\delta, \tau) = \begin{cases} 0 & \delta \in T_{\mathcal{B}}(0, \tau) \\ +\infty & \text{otherwise} \end{cases}.$$

Note that these functions are lower semicontinuous by construction. Chernoff regularity implies that (Geyer, 1994) $\sqrt{n}(\mathcal{B}(\cdot) - \beta(\cdot)) \to T_{\mathcal{B}}(0, \cdot)$, in other words that $w_n \xrightarrow{\text{epi}} w$. Also define $y_n(\delta, \tau) = -\delta^\top Z_{1n}(\tau) + \delta^\top H(\tau)\delta/2$ and $z(\delta, \tau) = -\delta^\top G(\tau) + \delta^\top H(\tau)\delta/2$, where the Gaussian process G is described in the statement of the theorem. The above functions are defined such that $Y_n = y_n + w_n$ and $Z = z + w$.

A calculation in Theorem 3 of Kato (2009) (specifically showing equation (14) in the proof) implies that $Z_{1n} \rightsquigarrow G$ in the space of bounded functions on $\mathcal{T}$. Use an almost sure representation (Dudley, 1985, Theorem 4.1) to construct a sequence $\{Z_{1n}^*\}$ such that $Z_{1n}^* \to G$ almost surely, where $Z_{1n}^* \sim Z_{1n}$ and $G^* \sim G$, and use this sequence to construct a sequence of functions $\{y_n^*\}$ and a function z^* that have analogous properties. Then $y_n^* \xrightarrow{\text{epi}} z^*$ almost surely by Theorem 7.46 of Rockafellar and Wets (1997), which implies $Y_n \rightsquigarrow Z$. The fact that $d_{\mathcal{F}}(Z_n, Y_n) \xrightarrow{p} 0$ and $Y_n \rightsquigarrow Z$ implies the result. ∎

Lemma A.1. *Let Z_{2n} be the function defined in (5) in the proof of Theorem 2.1. Then for each $\delta \in \sqrt{n}(\mathcal{B} - \beta_0)$, under Assumptions **A1**-**A7**,*

$$\sup_{\tau \in \mathcal{T}} \left| \mathrm{E}[Z_{2n}(\delta, \tau)] - \delta^\top H(\tau)\delta/2 \right| \xrightarrow{p} 0$$

and

$$\sup_{\tau \in \mathcal{T}} |Z_{2n}(\delta, \tau) - \mathrm{E}[Z_{2n}(\delta, \tau)]| \xrightarrow{p} 0.$$

Proof of Lemma A.1. The arguments here follow Kato (2009); here X is considered stochastic and there is no location-scale model imposed on the data. Once again let $U_i = Y_i - Q_{Y|X}(\tau|X_i) = Y_i - X_i^\top \beta_0(\tau)$. Consider the first assertion. Taking expectations conditional on X and using the independence between observations,

$$\begin{aligned} \mathrm{E}[Z_{2n}(\delta, \tau)|X] &= \frac{1}{n} \sum_i \delta^\top X_i \mathrm{E}\left[\int_0^1 \sqrt{n}\left(I(U_i \le (X_i^\top \delta/\sqrt{n}) \times s) - I(U_i \le 0) \right) \mathrm{d}s \middle| X_i \right] \\ &= \frac{1}{n} \sum_i \delta^\top X_i \left(\sqrt{n}\left(F_{Y|X}(Q_{Y|X}(\tau|X_i) + (X_i^\top \delta/\sqrt{n})s|X_i) - \tau \right) \right). \end{aligned}$$

For each term in the sum, the identity $\tau = F_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i)$ implies that

$$\sqrt{n}\left(F_{Y|X}(Q_{Y|X}(\tau|X_i)+(X_i^\top\delta/\sqrt{n})s|X_i)-\tau\right) = X_i^\top\delta\int_0^1 f_{Y|X}(Q_{Y|X}(\tau|X_i)+(X_i^\top\delta/\sqrt{n})s|X_i)\mathrm{d}s.$$

Note that the difference between this and $(X_i^\top\delta)f_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i)$ is uniformly negligible for large n

$$\begin{aligned}\left|X_i^\top\delta\int_0^1 f_{Y|X}(Q_{Y|X}(\tau|X_i)+(X_i^\top\delta/\sqrt{n})s|X_i)\mathrm{d}s - X_i^\top\delta f_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i)\right|& \\ \le \max_i|X_i^\top\delta|\int_0^1\left|f_{Y|X}(Q_{Y|X}(\tau|X_i)+(X_i^\top\delta/\sqrt{n})s|X_i)-f_{Y|X}(Q_{Y|X}(\tau|X_i))\right|\mathrm{d}s,&\end{aligned}$$

and assumption **A5** implies $\max_i|X_i^\top\delta| = o_P\left(n^{1/2}\right)$ (Angrist et al., 2006, part A.1.2), and

$$\max_i\sup_{\tau\in\mathcal{T}}\sup_{\delta:\|\delta\|\le K}\int_0^1|f_{Y|X}(Q_{Y|X}(\tau|X_i)+(X_i^\top\delta/\sqrt{n})s|X_i)-f_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i)|\mathrm{d}s = o_P(1)$$

for each $K>0$. Therefore for each δ,

$$\begin{aligned}&\frac{1}{n}\sum_i\delta^\top X_i\left(\sqrt{n}\left(F_{Y|X}(Q_{Y|X}(\tau|X_i)+(X_i^\top\delta/\sqrt{n})s|X_i)-\tau\right)\right)\\ &\qquad = \frac{1}{n}\sum_i\delta^\top X_i\left[\int_0^1 f_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i)s\mathrm{d}s\right]X_i^\top\delta + o_P(1).\end{aligned}$$

Letting $n\to\infty$ proves the first assertion.

Now turn to the second assertion. For any δ and each $\tau\in\mathcal{T}$

$$\begin{aligned}\mathrm{E}\left[(Z_{2n}(\tau,\delta)-\mathrm{E}[Z_{2n}(\tau,\delta)])^2\right] &\le \mathrm{E}\left[Z_{2n}^2(\tau,\delta)\right]\\ &= \frac{1}{n}\sum_i\mathrm{E}\left[\left(\delta^\top X_i\int_0^1 I(U_i\le(X_i^\top\delta/\sqrt{n})\times s)-I(U_i\le 0)\mathrm{d}s\right)^2\right]\\ &\le \frac{1}{n}\sum_i\mathrm{E}\left[(X_i^\top\delta)^2\right]\mathrm{P}\left\{|U_i|\le|X_i^\top\delta/\sqrt{n}|\right\}\end{aligned}$$

The probabilities in these terms are asymptotically negligible uniformly in τ by the uniform continuity of $f_{Y|X}$, and the bound on the moments of X implies that the difference above converges in mean square to zero; convergence in mean square implies convergence in probability. Next, to show that Z_{2n} is asymptotically equicontinuous in probability, rewrite U_i as $Y_i - Q_{Y|X}(\tau|X_i)$ to stress its dependence on τ and define for a bounded triangular array $\{\xi_{in}\}$

$$J_n(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n(X_i^\top\delta)\left(I(Y_i\le Q_{Y|X}(\tau|X_i)+X_i^\top\xi_{in})-I(Y_i\le Q_{Y|X}(\tau|X_i))\right).$$

Then some calculation shows that for any $\tau_1 \leq \tau \leq \tau_2$,

$$\begin{aligned}
&\mathrm{E}\left[(J_n(\tau)-J_n(\tau_1))^2 (J_n(\tau_2)-J_n(\tau))^2\right] \\
&\qquad \leq 3\mathrm{E}\left[\left(\frac{1}{n}\sum_i (X_i^\top \delta)^2 \left(F_{Y|X}\left(Q_{Y|X}(\tau_2|X)+X^\top \xi_{in}\right)-F_{Y|X}\left(Q_{Y|X}(\tau_1|X)+X^\top \xi_{in}\right)\right)\right)^2\right].
\end{aligned}$$

Because $\max_i \|X_i\| = o_P\left(n^{1/2}\right)$ and F is absolutely continuous, Theorem 13.5 of Billingsley (1999) shows that this implies asymptotic equicontinuity in probability of J_n, and uniform (in $\mathcal{T}$) convergence to zero, which also applies to its integral, Z_{2n}. ∎

Proof of Theorem 2.2. Theorem 2.1 shows that the sequence of finite-sample objective functions epi-converges to a limiting objective function. Use the almost sure representation theorem (Dudley, 1985) to choose Z_n^* and Z^* such that $Z_n^* \sim Z_n$ for each n, $Z^* \sim Z$ and $Z_n^* \overset{a.s.}{\to} Z^*$. Let $\tilde{\delta}_n(\tau) = \operatorname{argmin}_\delta Z_n^*(\delta,\tau)$ (with equality since Z^* is strictly convex) and $\delta^*(\tau) \in \operatorname{argmin}_\delta Z^*(\delta,\tau)$.

The objective functions are convex and level-bounded in δ locally uniformly in τ (Rockafellar and Wets, 1997, Definition 1.16) because they are convex and lower semicontinuous and $Z_n^*(0,\tau) = Z^*(0,\tau) = 0$. Therefore epi-convergence of Z_n to Z and Theorem 1.17 of Rockafellar and Wets (1997) shows that $\inf_\delta Z_n(\delta,\tau) \overset{a.s.}{\to} \inf_\delta Z(\delta,\tau)$ for each τ.

Theorem 7.41 of Rockafellar and Wets (1997) shows that because $Z_n^*(\cdot,\tau)$ is a convex lower semicontinuous function in δ, the solution $\delta_n^*(\tau)$ is a locally bounded and outer semicontinuous multifunction of τ. Corollary 7.43 of Rockafellar and Wets (1997) shows that δ^* is a single-valued, continuous function of τ because for each τ it is the minimizer of a strictly convex problem on the tangent set. Therefore for each $\eta > 0$,

$$\mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} d_{\mathcal{F}}(\limsup_n \delta_n^*(\tau), \delta^*(\tau)) > \eta\right\} = 0$$

which implies part (a). Part (b) follows from Proposition 3.1 of Geyer (1994) applied to each τ, epi-continuity in τ of the functions $\{Z_n^*\}$ to Z^* and the almost-sure representation. ∎

Proof of Lemma 2.3. Recall the definition $Z(\delta,\tau) = -\delta^\top G(\tau) + \delta^\top H(\tau)\delta/2$. Part (a) results from rewriting the objective function: because H is positive definite for all τ, Z can be rewritten as (recall $W(\tau) = H^{-1}(\tau)G(\tau)$)

$$Z(\delta,\tau) = \frac{1}{2}\delta^\top H(\tau)\delta - \delta^\top H(\tau)H^{-1}(\tau)G(\tau) \pm \frac{1}{2}W^\top(\tau)H(\tau)W(\tau) \tag{7}$$

$$= \frac{1}{2}(\delta - W(\tau))^\top H(\tau)(\delta - W(\tau)) - \frac{1}{2}W^\top(\tau)H(\tau)W(\tau). \tag{8}$$

Furthermore, the second term does not depend on δ, which implies part (b).

Finally, note that by part (b) $\tilde{\delta}(\tau)$ is a projection of $W(\tau)$ onto $\mathcal{C}$ using the norm $\|h\|_{H(\tau)} := (h^\top H(\tau)h)^{1/2}$. Therefore (in terms of this norm) $\tilde{\delta}(\tau)$ and $W(\tau) - \tilde{\delta}(\tau)$ are orthogonal — see, for example, Section 2 of Shapiro (1988). Rewrite (8) as the orthogonal decomposition

$$
\begin{aligned}
Z(\tilde{\delta}(\tau),\tau) &= \frac{1}{2}\left(\|\tilde{\delta}(\tau)-W(\tau)\|^2_{H(\tau)} - \|W(\tau)\|^2_{H(\tau)}\right) \\
&= -\frac{1}{2}\|\tilde{\delta}(\tau)\|^2_{H(\tau)}
\end{aligned} \tag{9}
$$

which is the statement in part (c). ■

Proof of Lemma 3.1. For any τ first note that the solution (A.11-A.12 in Appendix) implies

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\tilde{\lambda}_{1ni}(\tau) = \frac{1}{\sqrt{n}}\sum_{i\in h_1} X_i\tilde{\lambda}_{1ni} + \frac{1}{\sqrt{n}}\sum_{i\in\bar{h}_1} X_i\psi_\tau(Y_i - X_i^\top\tilde{\beta}(\tau)).
$$

Add and subtract $\frac{1}{\sqrt{n}}\sum_{i\in h_1} X_i\psi_\tau(Y_i - X_i^\top\tilde{\beta}(\tau))$ and note that $\psi_\tau(Y_i - X_i^\top\tilde{\beta}(\tau)) = \tau$ when $i\in h_1$ to find

$$
= \frac{1}{\sqrt{n}}\sum_{i\in h_1} X_i\left(\tilde{\lambda}_{1ni}(\tau)-\tau\right) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\psi_\tau(Y_i - X_i^\top\tilde{\beta}(\tau)).
$$

Because of assumption **A5** implies that $\max_i\|X_i\| = o_P\left(n^{1/2}\right)$, as detailed in part A.1.2 of Angrist et al. (2006), $-1\le\tilde{\lambda}_{1ni}-\tau\le 0$ and $0\le \mathrm{card}(h_1)\le p$,

$$
\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{\sqrt{n}}\sum_{i\in h_1} X_i\left(\tilde{\lambda}_{1ni}(\tau)-\tau\right)\right\| \le p\max_i\|X_i\|/\sqrt{n} = o_P(1)
$$

so that

$$
\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\tilde{\lambda}_{1ni}(\tau) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\psi_\tau(Y_i - X_i^\top\tilde{\beta}(\tau))\right\| = o_P(1). \tag{10}
$$

■

Proof of Lemma 3.2. Label the difference

$$
\Delta_n^S(\tau) := \frac{1}{\sqrt{n}}\sum_i X_i\psi_\tau(Y_i - X_i^\top\tilde{\beta}(\tau)) - \frac{1}{\sqrt{n}}\sum_i X_i\psi_\tau(Y_i - X_i^\top\beta_0(\tau)) + H(\tau)\sqrt{n}\left(\tilde{\beta}(\tau)-\beta_0(\tau)\right).
$$

Assumption **A1** implies the conditional quantiles of Y given $X = X_i$ are $Q_{Y|X}(\tau|X_i) = X_i^\top\beta_0(\tau)$. Define

$$
\tilde{g}_n(\beta,\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i\psi_\tau\left(Y_i - X_i^\top\beta\right), \quad g_n(\beta,\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i\left(\tau - F_{Y|X}(X_i^\top\beta|X_i)\right),
$$

and $\mathcal{G}_n(\alpha,\tau) = \sqrt{n}\,(\tilde{g}_n(\beta,\tau) - g_n(\beta,\tau))$.

Adding and subtracting $\sqrt{n} g_n(\tilde{\beta}(\tau), \tau)$ and $\sqrt{n} g_n(\beta_0(\tau), \tau)$, write $\Delta_n^S(\tau) = \xi_{1n}(\tau) + \xi_{2n}(\tau)$ where

$$\begin{aligned}
\xi_{1n}(\tau) &= \mathcal{G}_n(\tilde{\beta}(\tau), \tau) - \mathcal{G}_n(\beta_0(\tau), \tau), \\
\xi_{2n}(\tau) &= \sqrt{n}\left\{g_n(\tilde{\beta}(\tau), \tau) - g_n(\beta_0(\tau), \tau)\right\} + H(\tau)\sqrt{n}\left(\tilde{\beta}(\tau) - \beta_0(\tau)\right).
\end{aligned}$$

It will be shown that these two terms are uniformly asymptotically negligible.

Appendix A.2 of Kato (2009) (or similarly, the stochastic equicontinuity argument used in section A.1.2 of Angrist et al. (2006)) can be adapted in a straightforward manner to show that under Assumptions **A3**, **A5** and **A2**,

$$\sup_{\tau \in \mathcal{T}} \sup_{\beta \in \mathcal{B}(\tau): \|\beta\| < M} |\mathcal{G}_n(\beta, \tau) - \mathcal{G}_n(\beta_0(\tau), \tau)| \xrightarrow{p} 0$$

for each $M > 0$, that is, that $\sup_{\tau \in \mathcal{T}} \sup_{\beta \in \mathcal{B}(\tau): \|\beta\| < M} |\xi_{1n}(\tau)| = o_P(1)$ for each $M > 0$. Uniform consistency of $\tilde{\beta}$ then implies the first term is negligible.

For bounded $\delta \in \mathbb{R}^p$ such that $\beta_0(\tau) + \delta/\sqrt{n} \in \mathcal{B}(\tau)$,

$$\sqrt{n}\left(g_n(\beta_0(\tau) + \delta/\sqrt{n}, \tau) - g_n(\beta_0(\tau), \tau)\right) = \frac{-1}{\sqrt{n}} \sum_{i=1}^n X_i X_i^\top \delta/\sqrt{n} \int_0^1 f_{Y|X}\left(Q_{Y|X}(\tau|X_i) + s \times X_i^\top \delta/\sqrt{n} | X_i\right) \mathrm{d}s$$

and the gradient of g_n with respect to β evaluated at $\beta_0(\tau)$ is

$$\nabla_\beta g_n(\beta, \tau)|_{\beta = \beta_0(\tau)} = \frac{-1}{n} \sum_{i=1}^n f_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i) X_i X_i^\top := -H_n^0(\tau). \tag{11}$$

Then for any δ such that $\beta_0(\tau) + \delta/\sqrt{n} \in \mathcal{B}(\tau)$, $\|\delta\| < M$,

$$\begin{aligned}
&\left\|\sqrt{n}\left(g_n(\beta_0(\tau) + \delta/\sqrt{n}, \tau) - g_n(\beta_0(\tau), \tau)\right) + H_n^0(\tau)\delta\right\| = \\
&\qquad \left\|\frac{-1}{\sqrt{n}} \sum_{i=1}^n X_i X_i^\top \delta/\sqrt{n} \times \int_0^1 f_{Y|X}\left(Q_{Y|X}(\tau|X_i) + s X_i^\top \delta/\sqrt{n} | X_i\right) - f_{Y|X}(Q_{Y|X}(\tau|X_i)|X_i) \mathrm{d}s\right\|.
\end{aligned}$$

Then Assumptions **A2** and **A5** imply for each $M > 0$,

$$\sup_{\tau \in \mathcal{T}} \sup_{\beta \in \mathcal{B}(\tau): \|\beta\| < M} \left|\sqrt{n}(g_n(\beta, \tau) - g_n(\beta_0(\tau), \tau)) + H_n^0(\tau)\sqrt{n}(\beta - \beta_0(\tau))\right| = o_P\left(n^{-1/2}\right).$$

That is, $\sup_{\tau \in \mathcal{T}} \sup_{\beta \in \mathcal{B}(\tau): \|\beta\| < M} |\xi_{2n}(\tau)| = o_P(1)$ for each $M > 0$.

Let $\epsilon > 0$, and choose an M such that

$$\limsup_{n \to \infty} \mathrm{P}\left\{\sup_{\tau \in \mathcal{T}} \|\beta(\tau)\| > M\right\} < \epsilon.$$

It is possible to choose such an M because Theorem 2.2 shows that $\tilde{\beta}$ is uniformly $\sqrt{n}$-consistent over

$\mathcal{T}$. This all implies that for any $\gamma \geq 0$,

$$\limsup_{n\to\infty} \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \|\Delta_n^S(\tau)\| > \gamma\right\} \leq \limsup_{n\to\infty} \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \sup_{\beta\in\mathcal{B}(\tau):\|\beta\|<M} \|\Delta_n^S(\tau)\| > \gamma\right\} + \mathrm{P}\left\{\sup_{\tau\in\mathcal{T}} \|\beta_0(\tau)\| > M\right\}$$
$$\leq \limsup_{n\to\infty} \sum_{j=1}^{2} \sup_{\tau\in\mathcal{T}} \sup_{\beta\in\mathcal{B}(\tau):\|\beta\|<M} \left|\xi_{jn}(\tau)\right| + \epsilon.$$

Since ϵ is arbitrary we have that

$$\sup_{\tau\in\mathcal{T}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \psi_\tau(Y_i - X_i^\top \tilde{\beta}(\tau)) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \psi_\tau(Y_i - X_i^\top \beta_0(\tau)) + H_n^0(\tau)\sqrt{n}\left(\tilde{\beta}(\tau) - \beta_0(\tau)\right) \right| = o_P(1).$$

Because H_n^0 converges in probability to H uniformly in τ this implies the result. ∎

Proof of Theorem 4.1. The statement in part (a) is Corollary 1 of Koenker and Machado (1999, p. 1298); its proof is unchanged by restrictions on the parameter vector.

Note that the hypothetical parameter spaces can be rewritten in terms of δ under the null hypothesis H_0^A. Under H_0^A, the parameter space $\{\beta \in \mathbb{R}^p : R\beta = r\}$ is equivalent to $\{\delta \in \mathbb{R}^p : R\delta = \mathbf{0}_q\}$. Similarly, under H_0^A (not H_0^B), $\{\beta \in \mathbb{R}^p : R\beta \geq r\}$ is equivalent to $\{\delta \in \mathbb{R}^p : R\delta \geq \mathbf{0}_q\}$.

For each τ, $L_n^A(\tau)$ is a scaled version of the difference process

$$\bar{V}_n(\tau) - \tilde{V}_n(\tau) = Z_n(\bar{\delta}_n(\tau), \tau) - Z_n(\tilde{\delta}_n(\tau), \tau).$$

For each τ rewrite this as (Koenker and Bassett, 1982)

$$\begin{aligned} Z_n(\bar{\delta}_n(\tau), \tau) - Z_n(\tilde{\delta}_n(\tau), \tau) &= Z_n(\bar{\delta}_n(\tau), \tau) - Z(\bar{\delta}_n(\tau), \tau) + Z(\bar{\delta}_n(\tau), \tau) - Z(\bar{\delta}(\tau), \tau) \\ &\quad + Z(\bar{\delta}(\tau), \tau) - Z(\tilde{\delta}(\tau), \tau) \\ &\quad + Z(\tilde{\delta}(\tau), \tau) - Z(\tilde{\delta}_n(\tau), \tau) + Z(\tilde{\delta}_n(\tau), \tau) - Z_n(\tilde{\delta}_n(\tau), \tau) \\ &= Z(\bar{\delta}(\tau), \tau) - Z(\tilde{\delta}(\tau), \tau) + R(\tau) \end{aligned} \tag{12}$$

where by Theorems 2.1 and 2.2, $\sup_\tau \|R(\tau)\| = o_P(1)$. Note that convergence here is of bounded functions and the value functions are almost surely bounded by part (b) of Theorem 2.2, so that this is equivalent to usual weak convergence in the space of bounded functions. This and the continuous mapping theorem establish part (b).

To characterize the limiting distributions in part (b) note that under assumption **B1** with the known scale of the error distribution normalized to 1, the variance of $W(\tau)$ is the simpler $\Sigma_0(\tau)$. Assume H_0^A and rewrite (12) using Lemma 2.3, part (a), and scale it to find an asymptotic expression for the marginal distributions:

$$L_n^A(\tau) \rightsquigarrow \min_{\delta:R\delta=\mathbf{0}} (\delta - W(\tau))^\top \Sigma_0^{-1}(\tau)(\delta - W(\tau)) - \min_{\delta:R\delta\geq\mathbf{0}} (\delta - W(\tau))^\top \Sigma_0^{-1}(\tau)(\delta - W(\tau)). \tag{13}$$

Corollary 3.7.2 of Silvapulle and Sen (2005) states that the marginal distributions follow $\bar{\chi}^2$ distributions with the given weights for each τ, that is, that the process tends to a $\bar{Q}_A^2$ process.

To show part (c), rewrite (13) using the minimizers to each problem as

$$L_n^A(\tau) \rightsquigarrow \left(\bar{\delta}(\tau) - W(\tau)\right)^\top \Sigma_0^{-1}(\tau)\left(\bar{\delta}(\tau) - W(\tau)\right) - \left(\tilde{\delta}(\tau) - W(\tau)\right)^\top \Sigma_0^{-1}(\tau)\left(\tilde{\delta}(\tau) - W(\tau)\right).$$

Adding and subtracting $W^\top(\tau)\Sigma_0^{-1}W(\tau)$, note that because each asymptotic solution is a projection onto the subspaces $\mathcal{C}$ or $\mathcal{M}$ defined in the statement of the theorem, a Pythagorean theorem in the norm $\|v\|_{\Sigma_0^{-1}(\tau)} = (v^\top \Sigma_0^{-1} v)^{1/2}$ holds for each term, for example, $\|W(\tau)\|^2_{\Sigma_0^{-1}(\tau)} = \|\bar{\delta}(\tau) - W(\tau)\|^2_{\Sigma_0^{-1}(\tau)} + \|\bar{\delta}(\tau)\|^2_{\Sigma_0^{-1}(\tau)}$. Substituting in the appropriate terms and rearranging results in part (c).

To prove part (d), it is convenient to define

$$\begin{aligned} q(\delta, \tau) &= (\delta - W(\tau))^\top \Sigma_0^{-1}(\tau)(\delta - W(\tau)) \\ q_2(\delta_2, \tau) &= (\delta_2 - W_2(\tau))^\top \left(R\Sigma_0(\tau)R^\top\right)^{-1}(\delta_2 - W_2(\tau)) \end{aligned}$$

where R takes the special form specified in this part. These are nearly the same as (A.3) and (A.4) in the appendix except that they use Σ_0^{-1} in the place of H — under the assumption of homoskedasticity, these two matrices are scalar multiples of one another. Add and subtract a term to write the right-hand side of (13) as

$$\min_{\delta : R\delta = \mathbf{0}} q(\delta, \tau) - W^\top(\tau)\Sigma_0^{-1}(\tau)W(\tau) - \min_{\delta : R\delta \geq \mathbf{0}} q(\delta, \tau) + W^\top(\tau)\Sigma_0^{-1}(\tau)W(\tau).$$

The assumption about the decomposition of R made in this part is equivalent to condition **S** in the Appendix, so Theorem 1.1 of the appendix can be applied. Using part (a) of Theorem 1.1 in the appendix, with Σ_0^{-1} replacing H, this is equal to

$$\begin{aligned} &\min_{\delta : R\delta = \mathbf{0}} q_2(\delta, \tau) - W_2^\top(\tau)\left(R\Sigma_0(\tau)R^\top\right)^{-1} W_2(\tau) - G_1^\top(\tau)\left((I-R)\Sigma_0(\tau)(I-R)^\top\right)G_1(\tau) \\ &\qquad - \min_{\delta : R\delta \geq \mathbf{0}} q_2(\delta, \tau) + W_2^\top(\tau)\left(R\Sigma_0(\tau)R^\top\right)^{-1} W_2(\tau) + G_1^\top(\tau)\left((I-R)\Sigma_0(\tau)(I-R)\right)G_1(\tau). \end{aligned}$$

Cancel the parts that involve G_1 and, using the definition of q_2 and a Pythagorean theorem in $\|\cdot\|_{\Sigma_0^{-1}(\tau)}$ to rewrite this as the expression in part (d).

Turning to part (e), the analysis of type B processes is nearly identical to that of type A processes. Making the same series of manipulations one finds that the process converges to some stochastic process, which depends on the true parameter value so remains unknown because the parameter is generally restricted to a cone with nonempty interior. Under the hypothesis H_0^A, Corollary 3.8.3 of Silvapulle and Sen (2005) asserts that the asymptotic marginal distributions are $\bar{\chi}^2$ distributions with the weights stated in the description of the $\bar{Q}_B^2$ process.

Part (f) is shown in the same way as part (c) except that $\hat{\delta}(\cdot) = W(\cdot)$. ∎

Proof of Theorem 4.2. The continuous mapping theorem and Theorem 2.2 imply part (a). Suppose that

H_0^A is true and consider the form of this τ^{th} marginal difference in asymptotic quadratic minimization problems:

$$\Delta(\tau) = \min_{\delta:R\delta=\mathbf{0}} (\delta - W(\tau))^\top \Sigma^{-1}(\tau)(\delta - W(\tau)) - \min_{\delta:R\delta\geq\mathbf{0}} (\delta - W(\tau))^\top \Sigma^{-1}(\tau)(\delta - W(\tau)). \tag{14}$$

Because $\mathcal{M}$ is a linear subspace of $\mathcal{C}$, applying equation (3.7) of Shapiro (1988) shows that this follows a $\bar{\chi}^2$ distribution with weights given in the definition of the $\bar{Q}_A^2$ process. Now it remains to show that this random variable has the same distribution as the marginal asymptotic distribution of the type A Wald process.

Divide the parameter space into directions parallel to the constraints and orthogonal to them in the norm $\|v\|_{\Sigma^{-1}(\tau)} = (v^\top \Sigma^{-1}(\tau) v)^{1/2}$: following Amemiya (1985, Section 1.4.2), define another matrix $S \in \mathbb{R}^{s\times p}$ such that $M = [S^\top, R^\top]^\top$ is nonsingular and $S\Sigma(\tau)R^\top = \mathbf{0}_{s\times q}$. Then the quadratic forms in (14) can be rewritten. For example,

$$\begin{aligned}(\delta - W(\tau))^\top \Sigma^{-1}(\tau)(\delta - W(\tau)) &= (\delta - W(\tau))^\top M^\top \left(M^\top\right)^{-1} \Sigma^{-1}(\tau) M^{-1} M(\delta - W(\tau)) \\ &= (\delta - W(\tau))^\top M^\top \left(M\Sigma(\tau)M^\top\right)^{-1} M(\delta - W(\tau)) \\ &= (\delta - W(\tau))^\top M^\top \begin{bmatrix} (S\Sigma(\tau)S^\top)^{-1} & \mathbf{0}_{s\times q} \\ \mathbf{0}_{q\times s} & (R\Sigma(\tau)R^\top)^{-1} \end{bmatrix} M(\delta - W(\tau)) \\ &= (\delta - W(\tau))^\top S^\top \left(S\Sigma(\tau)S^\top\right)^{-1} S(\delta - W(\tau)) \\ &\quad + (\delta - W(\tau))^\top R^\top \left(R\Sigma(\tau)R^\top\right)^{-1} R(\delta - W(\tau)).\end{aligned}$$

When minimizing subject to the constraints $R\delta \geq \mathbf{0}$, it is always possible to set the first term equal to zero because it is orthogonal to the constraints. This means that the only (possibly) nonzero element is the final term involving R, for example, that

$$\min_{\delta:R\delta\geq\mathbf{0}} (\delta - W(\tau))^\top \Sigma^{-1}(\tau)(\delta - W(\tau)) = \min_{\delta:R\delta\geq\mathbf{0}} (\delta - W(\tau))^\top R^\top (R\Sigma(\tau)R^\top)^{-1} R(\delta - W(\tau)).$$

Applying these calculations to both halves of the right-hand side of (14), we have that

$$\Delta(\tau) = \left(\bar{\delta}(\tau) - \tilde{\delta}(\tau)\right)^\top R^\top (R\Sigma(\tau)R^\top)^{-1} R\left(\bar{\delta}(\tau) - \tilde{\delta}(\tau)\right) \sim \bar{\chi}^2,$$

with the weights given in the definition of the $\bar{Q}_A^2$ process. To find the marginal distributions of the asymptotic type B process asserted in part (b), the same algebraic manipulations as above can be applied, and combined with (3.1)-(3.6) of Shapiro (1988). Under H_0^B this follows some form of $\bar{Q}^2$ distribution, but under H_0^A this difference has the distribution of the $\bar{Q}_B^2$ process. ■

Proof of Theorem 4.3. Working with the type A process, manipulate $T_n^A(\tau)$ as follows: using Lemma 3.1

and Lemma 3.2, write the difference

$$\begin{aligned}\bar{S}_n(\tau) - \tilde{S}_n(\tau) &= -H_n^{-1}(\tau)H(\tau)\sqrt{n}(\bar{\beta}(\tau) - \beta_0(\tau)) + H_n^{-1}(\tau)H(\tau)\sqrt{n}(\tilde{\beta}(\tau) - \beta_0(\tau)) \\ &= \sqrt{n}\left(\tilde{\beta}(\tau) - \bar{\beta}(\tau)\right) + o_P(1),\end{aligned}$$

uniformly in τ, using the fact that $H_n(\cdot)$ is a uniformly consistent estimate of $H(\cdot)$ and nonsingular. Substitute this into $T_n^A(\tau)$, and the continuous mapping theorem implies the result. The type B process uses the same steps and the fact that $\hat{S}_n(\cdot) \equiv \mathbf{0}_p$. ■

References

J. Abrevaya. Isotonic quantile regression: Asymptotics and bootstrap. *Sankhyā*, 67:1–13, 2005.

T. Amemiya. *Advanced Econometrics*. Harvard University Press, Cambridge, 1985.

D.W.K. Andrews. Estimation when a parameter is on a boundary. *Econometrica*, 67:1341–1383, 1999.

D.W.K. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69:683–734, 2001.

D.W.K. Andrews and X. Shi. Inference based on many conditional moment inequalities. *Journal of Econometrics*, 196:275–287, 2017.

J. Angrist, V. Chernozhukov, and I. Fernández-Val. Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, 74:539–563, 2006.

R.E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk. *Statistical Inference under Order Restrictions*. Wiley, New York, 1972.

G. Beer. *Topologies on Closed and Closed Convex Sets*. Kluwer, Dordrecht, 1993.

P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 2 edition, 1999.

A. Bücher, J. Segers, and S. Volgushev. When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *The Annals of Statistics*, 42:1598–1634, 2014.

V. Chernozhukov, S. Lee, and A.M. Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81:667–737, 2013.

V. Chernozhukov, D. Chetverikov, and K. Kato. Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, forthcoming.

R. Dudley. An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. In A. Beck, R. Dudley, M. Hahn, J. Kuelbs, and M. Marcus, editors, *Probability in Banach Spaces V*, number 1153 in Lecture Notes in Mathematics, pages 141–178. Springer, Berlin, Heidelberg, 1985.

C.J. Geyer. On the asymptotics of constrained M-estimation. *The Annals of Statistics*, 22:1993–2010, 1994.

C. Gutenbrunner and J. Jurečková. Regression rankscores and regression quantiles. *The Annals of Statistics*, 20:305–330, 1992.

C. Gutenbrunner, J. Jurečková, R. Koenker, and S. Portnoy. Tests of linear hypotheses based on regression rankscores. *Journal of Nonparametric Statistics*, 2:307–331, 1993.

W. Hendricks and R. Koenker. Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87:58–68, 1992.

K. Kato. Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100: 1816–1829, 2009.

K. Knight. Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics*, 26:755–770, 1998.

R. Koenker. *Quantile Regression*. Cambridge University Press, Cambridge, 2005.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

R. Koenker and G. Bassett. Tests of linear hypotheses and l_1 estimation. *Econometrica*, 50:1577–1583, 1982.

R. Koenker and J.A.F. Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94:1296–1310, 1999.

R. Koenker and P. Ng. Inequality constrained quantile regression. *Sankhyā: The Indian Journal of Statistics*, 67:418–440, 2005.

S. Lee, K. Song, and Y.-J. Whang. Testing functional inequalities. *Journal of Econometrics*, 172:14–32, 2013.

O. Linton, K. Song, and Y.-J. Whang. An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154:186–202, 2010.

S. Portnoy and R. Koenker. The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statistical Science*, 12:279–300, 1997.

J. Powell. Censored regression quantiles. *Journal of Econometrics*, 32:143–155, 1986.

T. Robertson, F.T. Wright, and R.L. Dykstra. *Order Restricted Statistical Inference*. Wiley, New York, 1988.

R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*. Springer, Berlin, Heidelberg, 1997. Third printing 2009.

E. Ronchetti. Robust alternatives to the *F* test for the linear model. In W.P.G. Grossmann and W. Wertz, editors, *Probability and Statistical Inference*, pages 329–342. Reidel, Dordrecht, 1982.

J.O. Royset. Approximations and solution estimates in optimization. *Mathematical Programming*, forthcoming.

A. Shapiro. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, 56:49–62, 1988.

M.J. Silvapulle and P.K. Sen. *Constrained Statistical Inference*. Wiley, New York, 2005.

A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.

F.A. Wolak. The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica*, 59:981–995, 1991.

Appendix to “Asymptotic Inference for the Constrained Quantile Regression Process”

December 20, 2018

The first section describes asymptotic behavior of quantile regression coefficient estimates when only a subset of coefficients is constrained and the rest are unconstrained coefficients. Next there is a short discussion about obtaining dual solutions from the primal quantile regression coefficient estimate since at present no specialized software exists to construct such solutions automatically. Third, a section relates the rankscore statistic of Koenker and Machado (1999) to the rankscore statistics used in this article, which look rather different but are equivalent under the right conditions. Finally there is a section that offers two small simulation experiment examples for testing in a one-sample situation and in a simple regression model.

1 Subvector distributions

It may be of interest to know the distribution of parameter estimates when one subset of coefficient processes is constrained and the other is unconstrained. Assume that the covariates can be split into two groups that correspond to parameters that lie on respectively the boundary or the interior of the parameter set (as specified below). A more complex setting is analyzed in Andrews (2001).

Suppose that the coefficient vector β can be logically divided into two groups as $\beta_1 \in \mathbb{R}^s$, $\beta_2 \in \mathbb{R}^q$ such that $s + q = p$. β_1 represents unrestricted parameters and β_2 represents the vector of coefficients restricted by the null hypothesis to the boundary of the parameter set. Split the main components of the limit process into conformable pieces as

$$\delta(\cdot) = \begin{bmatrix} \delta_1(\cdot) \\ \delta_2(\cdot) \end{bmatrix} \quad H(\cdot) = \begin{bmatrix} H_{11}(\cdot) & H_{12}(\cdot) \\ H_{21}(\cdot) & H_{22}(\cdot) \end{bmatrix} \tag{A.1}$$

$$G(\cdot) = \begin{bmatrix} G_1(\cdot) \\ G_2(\cdot) \end{bmatrix} \quad W(\cdot) = \begin{bmatrix} W_1(\cdot) \\ W_2(\cdot) \end{bmatrix}. \tag{A.2}$$

Also define $\Xi = [\mathbf{0}_{q\times s}, I_q]$, $q \times p$ matrix that selects the β_2 coordinates from $\beta(\tau)$ and other conformably split vectors and matrices.

Assumption **S** implicitly defines which coefficients are restricted and which are not by specifying the form of the (marginal) tangent set.

S The asymptotic parameter space $\mathcal{C}$ can be subdivided as $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$, where $\mathcal{C}_1 = \mathbb{R}^s$ and $\mathcal{C}_2 \subset \mathbb{R}^q$ is a convex cone.

Finally, define two asymptotic quadratic objective functions like the one shown in part (b) of Lemma 2.3:

$$q(\delta,\tau) = (\delta - W(\tau))^\top H(\tau)(\delta - W(\tau)), \tag{A.3}$$

$$q_2(\delta_2,\tau) = (\delta_2 - W_2(\tau))^\top \left(\Xi H^{-1}(\tau)\Xi^\top\right)^{-1}(\delta_2 - W_2(\tau)). \tag{A.4}$$

The following theorem summarizes several properties of subvectors of $\tilde{\delta}(\tau)$. In particular, an analytic solution is can be found for parameters that lie in the interior of the parameter space, although they depend on the value of the parameters on the boundary.

Theorem 1.1. *Fix a value of $\tau \in \mathcal{T}$. Divide the solution vector $\tilde{\delta}(\tau)$ into two subvectors $\tilde{\delta}_1(\tau)$ and $\tilde{\delta}_2(\tau)$ corresponding to the definition in (A.1). Under assumptions **A3**, **A4**, **A6**, **A7** and **S**,*

(a) $W^\top(\tau)H(\tau)W(\tau) = G_1^\top(\tau)H_{11}^{-1}(\tau)G_1(\tau) + W_2^\top(\tau)\left(\Xi H^{-1}(\tau)\Xi^\top\right)^{-1}W_2(\tau)$.

(b) $\min_{\delta\in\mathcal{C}} q(\delta,\tau) = \min_{\delta_2\in\mathcal{C}_2} q_2(\delta_2,\tau)$

(c) $\tilde{\delta}_2(\tau) = \operatorname{argmin}_{\delta_2\in\mathcal{C}_2} q_2(\delta_2,\tau)$

(d) $\tilde{\delta}_1(\tau) = H_{11}^{-1}(\tau)W_1(\tau) - H_{11}^{-1}(\tau)H_{12}(\tau)\tilde{\delta}_2(\tau)$

Proof of Theorem 1.1. This proof follows the method developed in the proof of Theorem 4 of Andrews (1999) to divide the asymptotic quadratic form into parts that depend only on β_1 or β_2. First define

$$A(\tau) := \begin{bmatrix} -H_{11}^{-1}(\tau)H_{12}(\tau) \\ I_q \end{bmatrix} \in \mathbb{R}^{p\times q}.$$

Then define $P^\perp(\tau) := A(\tau)\Xi$ and $P(\tau) := I_p - P^\perp(\tau)$, that is,

$$P^\perp(\tau) = \begin{bmatrix} \mathbf{0}_{s\times s} & -H_{11}^{-1}(\tau)H_{12}(\tau) \\ \mathbf{0}_{q\times s} & I_q \end{bmatrix}, \quad P(\tau) = \begin{bmatrix} I_s & H_{11}^{-1}(\tau)H_{12}(\tau) \\ \mathbf{0}_{q\times s} & \mathbf{0}_{q\times q} \end{bmatrix}.$$

Then in terms of the norm $\|\cdot\|_{H(\tau)}$ defined for $x \in \mathbb{R}^p$ by $\|x\|_{H(\tau)} = (x^\top H(\tau)x)^{1/2}$, $P(\tau)$ projects vectors in $\mathbb{R}^p$ onto the subspace $L = \left\{x \in \mathbb{R}^p : x = [u^\top, \mathbf{0}_q^\top]^\top \text{ for } u \in \mathbb{R}^s\right\}$ and $P^\perp(\tau)$ projects vectors onto its orthogonal complement. That is, for any $x, y \in \mathbb{R}^p$, $(P^\perp(\tau)x)^\top H(\tau)(P(\tau)y) = 0$.

The following quadratic form in W can be split into parts corresponding to β_1 and β_2 using the orthogonal projections $P^\perp(\tau)$ and $P(\tau)$:

$$W^\top(\tau)H(\tau)W(\tau) = \left(P^\perp(\tau)W(\tau)\right)^\top H(\tau)\left(P^\perp(\tau)W(\tau)\right) + (P(\tau)W(\tau))^\top H(\tau)(P(\tau)W(\tau)). \tag{A.5}$$

Note that

$$A^\top H(\tau)A = \left(\Xi H^{-1}(\tau)\Xi\right)^{-1} \tag{A.6}$$

can be verified by algebra, and also that

$$P(\tau)H^{-1}(\tau) = \begin{bmatrix} H_{11}^{-1}(\tau) & \mathbf{0}_{s\times q} \\ \mathbf{0}_{q\times s} & \mathbf{0}_{q\times q} \end{bmatrix}.$$

The $(2,2)$ block of the above matrix is the result of direct calculation after writing $H^{-1}(\tau)$ as a partitioned inverse. These equivalencies imply that

$$P(\tau)W(\tau) = \begin{bmatrix} H_{11}^{-1}(\tau)G_1(\tau) \\ \mathbf{0}_q \end{bmatrix}. \tag{A.7}$$

Use (A.6) and (A.7) to rewrite (A.5) as

$$W^\top(\tau)H(\tau)W(\tau) = G_1^\top(\tau)H_{11}^{-1}(\tau)G_1(\tau) + W_2^\top(\tau)\left(\Xi H^{-1}(\tau)\Xi^\top\right)^{-1} W_2(\tau),$$

which is the decomposition in part (a).

Now define another quadratic objective function,

$$q_1(\delta_1, \delta_2, \tau) = \left(\delta_1 + H_{11}^{-1}(\tau)H_{12}(\tau)\delta_2 - H_{11}^{-1}(\tau)G_1(\tau)\right)^\top H_{11}(\tau)\times \\ \left(\delta_1 + H_{11}^{-1}(\tau)H_{12}(\tau)\delta_2 - H_{11}^{-1}(\tau)G_1(\tau)\right).$$

Using the fact that

$$P\delta(\tau) = \begin{bmatrix} \delta_1(\tau) + H_{11}^{-1}(\tau)H_{12}(\tau)\delta_2(\tau) \\ \mathbf{0}_q \end{bmatrix},$$

make the orthogonal decomposition

$$q(\delta,\tau) = \left(P^\perp(\tau)\delta - P^\perp(\tau)W(\tau)\right)^\top H(\tau)\left(P^\perp(\tau)\delta - P^\perp(\tau)W(\tau)\right) + \\ (P(\tau)\delta - P(\tau)W(\tau))^\top H(\tau)(P(\tau)\delta - P(\tau)W(\tau))$$

and (A.7) to rewrite $q(\delta,\tau) = q_1(\delta_1,\delta_2,\tau) + q_2(\delta_2,\tau)$. For any given value of the subvector δ_2, q_1 is a quadratic form that is minimized over $\mathcal{C}_1 = \mathbb{R}^s$, which implies that $\min_{\delta_1\in\mathcal{C}_1} q_1(\delta_1,\delta_2,\tau) = 0$. This implies part (b).

To show part (c), note that

$$\begin{aligned} 0 &\le q_2(\tilde{\delta}_2(\tau),\tau) - \min_{\delta_2\in\mathcal{C}_2} q_2(\delta_2,\tau) \\ &\le q_1(\tilde{\delta}_1(\tau),\tilde{\delta}_2(\tau),\tau) + q_2(\tilde{\delta}_2(\tau),\tau) - \min_{\delta_2\in\mathcal{C}_2} q_2(\delta_2,\tau) \\ &= q(\tilde{\delta}(\tau),\tau) - \min_{\delta\in\mathcal{C}} q(\delta,\tau) = 0, \end{aligned}$$

where the equality follows from part (b), the decomposition of the quadratic objective function and the definition of $\tilde{\delta}(\tau)$. This implies part (c).

Finally, since $q(\tilde{\delta}(\tau),\tau)=\min_{\delta\in\mathcal{C}} q(\delta,\tau)=\min_{\delta_2\in\mathcal{C}_2} q_2(\delta_2,\tau)$ and $q(\tilde{\delta}(\tau),\tau)=q_1(\tilde{\delta}_1(\tau),\tilde{\delta}_2(\tau),\tau)+q_2(\tilde{\delta}_2(\tau),\tau)$, we know that $q_1(\tilde{\delta}_1(\tau),\tilde{\delta}_2(\tau),\tau)=\min_{\delta_1\in\mathcal{C}_1} q_1(\delta_1,\tilde{\delta}_2(\tau),\tau)=0$, which implies the solution in part (d). ■

2 Recovering inequality constrained dual solutions

Putting (3) in canonical form (Boyd and Vandenberghe, 2004, p. 147) makes it easier to transition to the corresponding dual problem. The canonical primal problem is defined as

$$\min_x \left\{c^\top x : Ax-b\in T, x\in S\right\},$$

which has dual problem

$$\max_y \left\{b^\top y : c-A^\top y\in S^*, y\in T^*\right\}.$$

where S^* and T^* are dual spaces associated with S and T (Koenker and Ng, 2005, p. 420).

To rewrite the primal problem in canonical form, let

$$x=\begin{bmatrix}u\\ v\\ b\end{bmatrix},\ c=\begin{bmatrix}\tau\mathbf{1}_n\\ (1-\tau)\mathbf{1}_n\\ \mathbf{0}_p\end{bmatrix},\ A=\begin{bmatrix}I_n & -I_n & X\\ \mathbf{0}_{q\times n} & \mathbf{0}_{q\times n} & R\end{bmatrix},\ b=\begin{bmatrix}Y\\ r\end{bmatrix}$$

and $T=\mathbf{0}_n\times\mathbb{R}^q_+$ and $S=\mathbb{R}^{2n}_+\times\mathbb{R}^p$. Then the minimization problem can be written

$$\min_{u,v,b}\left\{\begin{bmatrix}\tau\mathbf{1}_n^\top & (1-\tau)\mathbf{1}_n^\top & \mathbf{0}_p^\top\end{bmatrix}\begin{bmatrix}u\\ v\\ b\end{bmatrix} : \begin{bmatrix}I_n & -I_n & X\\ \mathbf{0}_{q\times n} & \mathbf{0}_{q\times n} & R\end{bmatrix}\begin{bmatrix}u\\ v\\ b\end{bmatrix}-\begin{bmatrix}Y\\ r\end{bmatrix}\in\mathbf{0}_n\times\mathbb{R}^q_+,\ \begin{bmatrix}u\\ v\\ b\end{bmatrix}\in\mathbb{R}^{2n}_+\times\mathbb{R}^p\right\}.$$

This has the dual problem

$$\max_\lambda\left\{\begin{bmatrix}Y^\top & r^\top\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} : \begin{bmatrix}\tau\mathbf{1}_n\\ (1-\tau)\mathbf{1}_n\\ \mathbf{0}_p\end{bmatrix}-\begin{bmatrix}I_n & \mathbf{0}_{n\times q}\\ -I_n & \mathbf{0}_{n\times q}\\ X^\top & R^\top\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}\in\mathbb{R}^{2n}_+\times\mathbf{0}_p, \lambda\in\mathbb{R}^n\times R^q_+\right\}. \tag{A.8}$$

Let $\tilde{\lambda}=[\tilde{\lambda}_1^\top,\tilde{\lambda}_2^\top]^\top$ denote the solution to the dual problem (A.8). Using the definitions of $\tilde{Y}$ and $\tilde{X}$ made in the main text, the dual program (A.8) can be collapsed into the more compact expression

$$\max_\lambda\left\{\tilde{Y}^\top\lambda : \tilde{X}^\top\lambda=\mathbf{0}_{p+q}, \lambda\in[\tau-1,\tau]^n\times\mathbb{R}^q_+\right\}. \tag{A.9}$$

Strict duality ensures that once the optimum has been found, the dual value function is identical to the primal value function (Boyd and Vandenberghe, 2004, p. 226-227), that is,

$$\tilde{Y}^\top\tilde{\lambda}=\tau\mathbf{1}_n^\top\tilde{u}+(1-\tau)\mathbf{1}_n^\top\tilde{v} \tag{A.10}$$

where $\tilde{u}$ and $\tilde{v}$ are vectors of positive and negative residuals from the constrained quantile regression fit, respectively.

The individual dual variables can be recovered from the primal solution and the definitions of h_1 and h_2. To satisfy equation (A.10), first solve for all the coordinates in $\bar{h}$ by setting

$$\tilde{\lambda}(\bar{h}) = \{\tilde{\lambda}_i\}_{i \in \bar{h}} = \begin{cases} \tau & i \in \bar{h}_1,\ \tilde{u}_i > 0 \\ \tau - 1 & i \in \bar{h}_1,\ \tilde{v}_i > 0 \\ 0 & i \in \bar{h}_2. \end{cases} \tag{A.11}$$

After this mapping process there remain the p non-assigned elements in h. Solve for these final terms by using the other condition

$$\begin{aligned} \tilde{X}^\top \tilde{\lambda} = \mathbf{0}_p &\Leftrightarrow \tilde{X}^\top(\bar{h})\tilde{\lambda}(\bar{h}) + \tilde{X}^\top(h)\tilde{\lambda}(h) = \mathbf{0}_p \\ &\Leftrightarrow \tilde{\lambda}(h) = -(\tilde{X}^\top(h))^{-1}\tilde{X}^\top(\bar{h})\tilde{\lambda}(\bar{h}). \end{aligned} \tag{A.12}$$

Finally, a note on implementation: the R package `quantreg` (Koenker, 2017) can be used to solve for the primal solution, that is, the constrained quantile regression coefficients. By default the constrained estimator uses an interior point method to find solutions. This is not guaranteed to end exactly on a basic solution, and has a tendency to find solutions that lie between observations/constraints. An ad-hoc procedure for finding the basic solution is to choose the indices of the smallest p quantile residuals; more careful programs, however, have capabilities to identify basic solutions after running an interior point algorithm. See section 10.3.2.4 ("Basis Identification") of the Rmosek manual (MOSEK ApS, 2018).

3 Relationship between constrained rankscore statistic and Koenker and Machado (1999)

At a casual glance it may be difficult to recognize that the rankscore process proposed in the main text is related to the rankscore process defined in Koenker and Machado (1999). They were concerned with the hypothesis that the last q quantile regression coefficients are equal to zero against an unrestricted alternative. To rewrite their statistic, first define partitions of the $\bar{H}_n$ matrix

$$\bar{H}_{njk}(\tau) = \frac{1}{nh_n}\sum_{i=1}^{n} K((Y_i - X_i^\top \bar{\beta}(\tau))/h_n)X_{ji}X_{ki}^\top, \quad j,k \in \{1,2\}.$$

Then Koenker & Machado define (these terms are rewritten a little differently from the original to accommodate more general forms of heteroskedasticity than were considered in that paper)

$$\begin{aligned}
\hat{X}_2 &= X_1 \bar{H}_{n11}^{-1}(\tau)\bar{H}_{n12}(\tau) \\
M_{KM} &= \tau(1-\tau)(X_2 - \hat{X}_2)^\top (X_2 - \hat{X}_2)/n \qquad \text{(A.13)} \\
\bar{S}_{KM}(\tau) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{2i} - \hat{X}_{2i})\bar{\lambda}_{1ni}(\tau) \\
T_{KM}(\tau) &= \bar{S}_{KM}^\top(\tau) M_{KM}^{-1} \bar{S}_{KM}(\tau). \qquad \text{(A.14)}
\end{aligned}$$

To show that T_{KM} is comparable to the regression rankscore statistics considered here, set the matrix R to the selector matrix $\Xi = [\mathbf{0}_{q\times s}, I_q]$ used earlier (where $s+q=p$), which corresponds to the hypothesis H_0 : $\beta_2 = \mathbf{0}$, and recall that in general any linear restriction $R\beta = r$ in a linear model can be reparameterized to an equivalent test of a zero restriction (Davidson and MacKinnon, 1993, Section 1.3). The T_{KM} statistic implicitly uses the unrestricted alternative hypothesis because the score based on unrestricted estimation is identically zero. In other words, the score processes discussed in Koenker and Machado (1999) are equivalent to those here except for the form of the null and alternative hypotheses. In Lemma 3.1 the right-hand side of the equality is written as the rankscore statistics referred to in the main text.

Lemma 3.1. *Suppose β is estimated under the hypothesis $H_0 : \Xi\beta(\tau) = \mathbf{0}_q$. Consider the regression rankscore test statistic T_{KM} defined in (A.14) for testing this restriction against an unrestricted alternative. Then*

$$T_{KM}(\tau) = \bar{S}_n(\tau)^\top \Xi^\top \left(\Xi \bar{\Sigma}_n(\tau) \Xi^\top\right)^{-1} \Xi \bar{S}_n(\tau). \qquad \text{(A.15)}$$

Proof of Lemma 3.1. Partition $\bar{H}_n(\tau)$ into four conformable matrices — the upper left an $s \times s$ matrix $\bar{H}_{n11}$ and lower right a $q \times q$ matrix $\bar{H}_{n22}$ for example. In this proof, the partitions of $\bar{H}_n$ are referred to without bars to reduce notational clutter. Define

$$A_n(\tau) = \begin{bmatrix} -H_{n11}^{-1}(\tau)H_{n12}(\tau) \\ I_q \end{bmatrix} \in \mathbb{R}^{p\times q}.$$

$A_n(\tau)$ can be used to rewrite the $\hat{X}_2$ terms used in Koenker and Machado (1999). In particular, $\hat{X}_2 = X_1 H_{n11}^{-1}(\tau)H_{n12}(\tau)$ so that it can be verified that

$$X_2 - \hat{X}_2 = XA_n(\tau).$$

This also implies $M_{KM} = \tau(1-\tau)A_n^\top(\tau)D_n A_n(\tau)$ where M_{KM} was defined in (A.13). Finally T_{KM} can be rewritten

$$T_{KM}(\tau) = \bar{\lambda}_n^\top(\tau) X A_n(\tau) \left(\tau(1-\tau)A_n^\top(\tau) D_n A_n(\tau)\right)^{-1} A_n^\top(\tau) X^\top \bar{\lambda}_n(\tau).$$

Now consider the formula on the right-hand side of (A.15): using a partitioned matrix formula it can be verified that

$$\begin{aligned}\Xi H_n^{-1}(\tau)X^\top &= \left(H_{n22}(\tau) - H_{n21}(\tau)H_{n11}^{-1}(\tau)H_{n12}(\tau)\right)^{-1}\left(X_2^\top - H_{n21}(\tau)H_{n11}^{-1}(\tau)X_1^\top\right) \\ &= H_n^{22}(\tau)A_n^\top(\tau)X^\top,\end{aligned}$$

using H_n^{22} to denote the $(2,2)$ block of H_n^{-1}. Plugging this into the right-hand side of expression (A.15), the result holds if

$$\begin{aligned}\left(A_n^\top(\tau)D_nA_n(\tau)\right)^{-1} &= H_n^{22}(\tau)\left(\Xi H_n^{-1}(\tau)D_nH_n^{-1}(\tau)\Xi^\top\right)^{-1}H_n^{22}(\tau) \\ \Leftrightarrow A_n^\top(\tau)D_nA_n(\tau) &= \left(H_n^{22}(\tau)\right)^{-1}\left(\Xi H_n^{-1}(\tau)D_nH_n^{-1}(\tau)\Xi^\top\right)\left(H_n^{22}(\tau)\right)^{-1} \\ \Leftrightarrow H_n^{22}(\tau)A_n^\top(\tau)D_nA_n(\tau)H_n^{22}(\tau) &= \Xi H_n^{-1}(\tau)D_nH_n^{-1}(\tau)\Xi^\top.\end{aligned}$$

This last equivalency is true because, for example $H_n^{22}(\tau)A_n(\tau)^\top = \Xi H_n^{-1}(\tau)$, which can be verified using a partitioned matrix formula. ■

4 Examples

Example 1. Here is a simple illustration of the test statistics discussed in the main text. Suppose that $X = 1$, so that the quantile regression estimator is $Q_Y(\tau) = \alpha(\tau)$, and we would like to test whether the (unconditional) distribution of Y dominates the standard normal distribution at first order. In terms of quantile functions that is the condition that $\alpha(\tau) \geq \Phi^{-1}(\tau)$, where Φ^{-1} is the standard normal quantile function. Typically in the econometrics literature a null of dominance is used, and the alternative is that a point $\tau \in \mathcal{T}$ exists where the distribution of Y does not dominate (Linton et al., 2010). Constrained quantile regression estimates make it straightforward to design a test for these hypotheses. Consider tests of

$$\begin{aligned}&H_0 : Q_Y(\tau) \geq \Phi^{-1}(\tau) \text{ for all } \tau \in \mathcal{T} \\ &H_1 : \text{there is some } \tau_0 \in \mathcal{T} \text{ such that } Q_Y(\tau_0) < \Phi^{-1}(\tau_0).\end{aligned}$$

The two hypotheses make this a type B problem. The model is estimated two ways; it is estimated once without constraints and once subject to the null constraint that $\alpha(\tau) \geq \Phi^{-1}(\tau)$. Evidence of a significantly positive effect at some quantile level would be indicated by a large value of one of the test statistics.

The problem is marginally one-dimensional, so constrained estimates are simply $\tilde{\alpha}(\cdot) = \hat{\alpha}(\cdot) \vee \Phi^{-1}(\cdot)$, where $\hat{\alpha}$ is the unconstrained quantile estimate. Then $\tilde{\delta}_n(\cdot) = \sqrt{n}(\tilde{\alpha}(\cdot) - \Phi^{-1}(\cdot))$ and its limit belong to the marginally polyhedral tangent set $T_{\mathcal{B}}(\beta_0) = \mathbb{R}_+ \times \mathcal{T}$.

Using Theorem 2.1, $\tilde{\delta}$ is marginally characterized by the simple problem $\tilde{\delta}(\tau) = \operatorname{argmin}_{\delta \in \mathbb{R}_+} -\delta B(\tau) + \delta^2 \phi(\Phi^{-1}(\tau))/2$, where ϕ is the standard normal density and B is a standard Brownian bridge. That is, $\tilde{\delta}(\cdot) = \phi^{-1}(\Phi^{-1}(\cdot))(B(\cdot) \vee 0)$. Lemma 4.1 part (f) shows that, under the least favorable null, for each τ the likelihood ratio process is marginally distributed as $\mathscr{L}^B(\tau) \sim (B(\tau) \wedge 0)^2/(\tau(1-\tau))$, which can be easily simulated.

The two most popular functionals for measuring the distance between inference processes such as these are the supremum- and L^2 norm metrics: for any function $f : \mathcal{T} \to \mathbb{R}$, let the sup-norm statistic derived used here is $\|f\|_\infty := \sup_{\tau \in \mathcal{T}} f(\tau)$, and $\|f\|_2 := \int_{\mathcal{T}} (f(\tau) \vee 0)^2 \mathrm{d}\tau$. To make a statistical decision and to inspect the performance of the asymptotic theory, p-values are convenient. Given an observed inference process and derived statistic $\tilde{T}$ using one of the norms described above, simulate the asymptotic limit process n_{sim} times, obtain $T_k^* = \|\mathscr{L}_k^B\|$ for $k = 1, \ldots n_{sim}$ and calculate $p^* = n_{sim}^{-1} \sum_{i=1}^{n_sim} I(T_k^* > \tilde{T})$. Under the null hypothesis, p^* should be uniformly distributed.

To produce the p-value plots in Figures 1 and 2, samples of size 100, 400 and 1000 were generated from a standard normal distribution and regressed on an intercept, either without constraint (so they are the sample quantiles) or under the dominance constraint. The inference processes are discretized — quantile regression coefficients are evaluated for $\tau \in \{0.05, 0.06, \ldots, 0.95\}$ and the norms are applied over this grid. The asymptotic process is also simulated on the same discretized grid. 1000 processes were estimated for each sample size and 1000 simulated processes were used to construct a reference distribution.

Figure 1 shows p-value plots derived from sup-norm statistics. In left-side panels, CDFs of p-values derived from an inference process are plotted against the theoretical uniform target CDF. On the right, the difference between empirical and theoretical CDFs are plotted. Figure 1 reveals a close correspondence between the p-values generated, even when the sample size is fairly small. A slight improvement can be seen when increasing the sample size.

The picture is similar for the L^2 statistics. Figure 2 shows p-value plots for these statistics. The same sample sizes and statistics are shown as in Figure 1, except that a different functional is used to evaluate the evidence against the null contained in the inference processes. Figure 2 looks very similar to the supremum norm results.

Example 2. Suppose that the random variable D denotes a treatment that is assigned randomly, and we can be confident that the treatment does not negatively affect agents. Then a plausible model might be

$$Q_{Y|X,D}(\tau|X, D) = X^\top \beta_1(\tau) + \beta_2(\tau) D, \tag{A.16}$$

estimated subject to the maintained hypothesis that $\beta \in \mathbb{R}^{p-1} \times \mathbb{R}_+$ for all τ (it is assumed that X contains an intercept). Suppose we wish to test

$$\begin{aligned} H_0 &: \beta_2(\tau) = 0 \text{ for all } \tau \in \mathcal{T} \\ H_1 &: \beta_2(\tau_0) > 0 \text{ for some } \tau_0 \in \mathcal{T}. \end{aligned}$$

Then a test can be conducted that uses equality constrained and inequality constrained estimates. The use of these particular estimates makes this a type A problem. Once again, evidence of a significantly positive effect at some quantile level would be indicated by a large value of one of the test statistics. For this problem $R = [\mathbf{0}_{p-1}^\top, 1]$ and $r = 0$. It is not assumed that the data are identically distributed, so only Wald and regression rankscore processes are considered here.

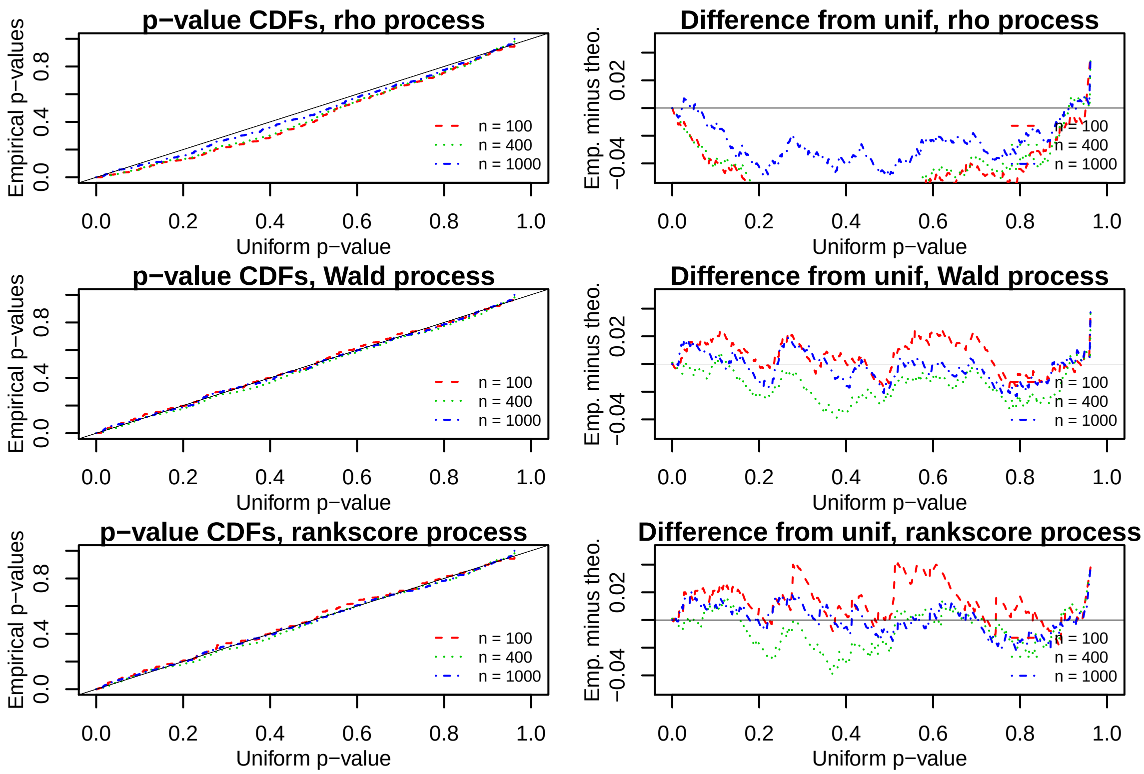

Figure 1: p-value plots for sup-norm statistics derived from the type-B inference processes described in the text for the one-sample data generating process described in the first example. The left panels show the CDFs of the p-values against the uniform CDF. The right panels show differences between the empirical p-value CDFs and the uniform CDF (vertical scale from -5% to +5% so all plots have the same scale). 1000 processes were estimated, and 1000 limit processes were simulated to serve as a reference distribution and calculate simulation based p-values.

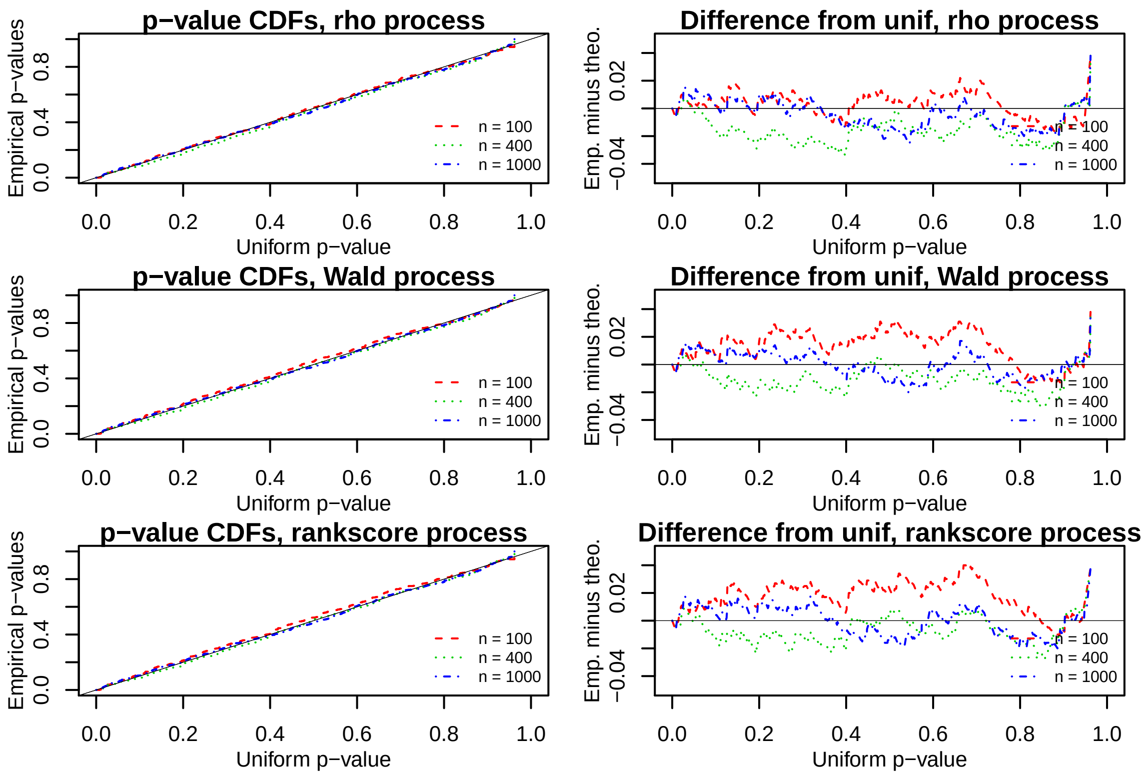

Figure 2: p-value plots for L^2 norm statistics in the one-sample example. The simulation details are the same as for the sup-norm statistics with the one-sample data generating process.

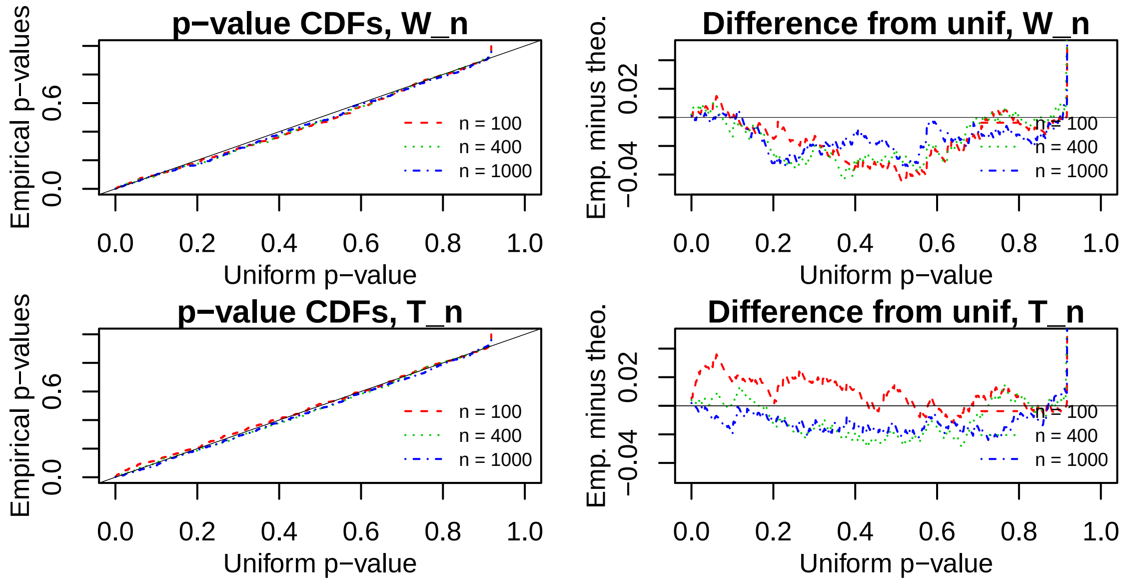

Figure 3: p-value plots for sup-norm statistics derived from the type A inference processes for detecting a positive treatment effect in the regression data generating process example. The left panels show the CDFs of the p-values against the uniform CDF. The right panels show differences between the empirical p-value CDFs and the uniform CDF (vertical scale from -5% to +5% so all plots have the same scale). 1000 processes were estimated, and 1000 limit processes were simulated to serve as a reference distribution and calculate simulation based p-values.

The suggested matrix Σ_n discussed above is used, using the default settings available in the R package quantreg (Koenker, 2017) — that is, from the standard error option `'ker'`, involving a Gaussian kernel with a robust estimate of the scale of the distribution. The limit process must be simulated using the estimate Σ_n since the weights in the $\bar{Q}^2$ process generally depend on Σ, and so the process does not have a pivotal distribution. For each process in the example below, a p-value is simulated using Σ_n and 1000 simulated Gaussian processes.

In the simulations for this example p was set to 4 and all regressors were generated as independent standard normal random variables. For each of 1000 simulation runs, Σ was estimated and a p-value was generated from 1000 simulated asymptotic statistics (either sup-norm or L^2-norm) that depend on Σ_n. To simulate the asymptotic process, let $q(\delta,\tau) = (\delta - W(\tau))^\top \Sigma^{-1}(\tau)(\delta - W(\tau))$ and note that asymptotically, both the Wald and score processes have the same marginal distributions as the difference $\min_{\delta:\delta_2=0} q(\delta,\tau) - \min_{\delta:\delta_2\geq 0} q(\delta,\tau)$, where q was defined in (A.3). Reduce this to the difference $(W_2(\tau)\wedge 0)^2/\Sigma_{22}(\tau)$, which can be simulated given an estimate of Σ_{22}.

Figure 3 reveals that both processes produce very similar results, and Figure 4 reveals that the L^2 statistic behaves much as the supremum norm statistic. These statistics are most likely so similar because of the relatively simple data generating process used for simulations, and may produce different results under heteroskedastic designs, for example.

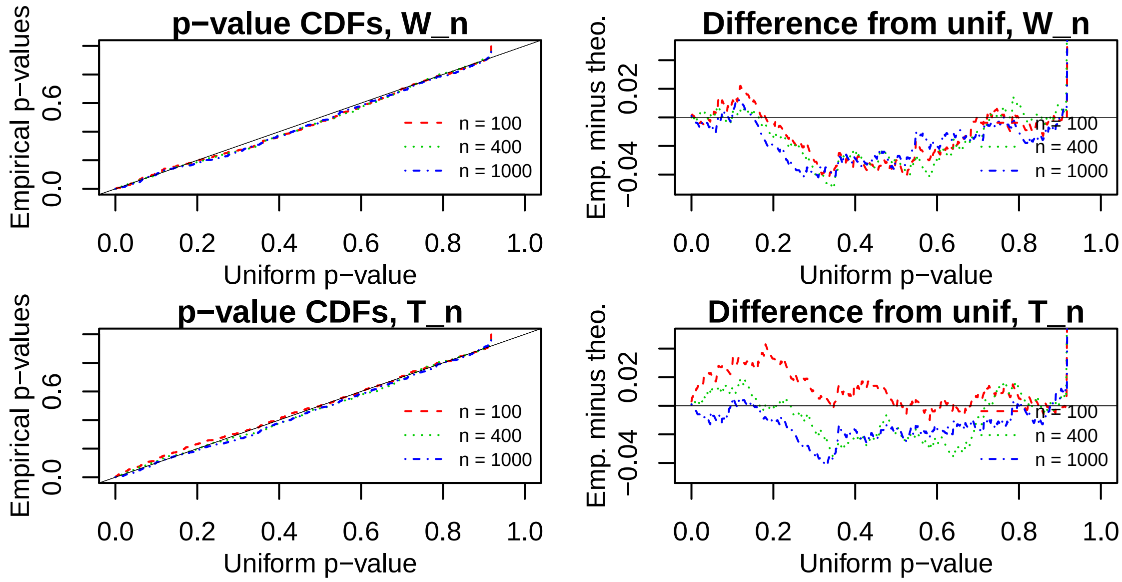

Figure 4: p-value plots for L^2 norm statistics in the regression example. The simulation details are the same as for the sup-norm statistics with the regression data generating process.

References

D.W.K. Andrews. Estimation when a parameter is on a boundary. *Econometrica*, 67:1341–1383, 1999.

D.W.K. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69:683–734, 2001.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

R. Davidson and J.G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, New York, 1993.

R. Koenker and J.A.F. Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94:1296–1310, 1999.

R. Koenker and P. Ng. Inequality constrained quantile regression. *Sankhyā: The Indian Journal of Statistics*, 67:418–440, 2005.

Roger Koenker. *quantreg: Quantile Regression*, 2017. URL `https://CRAN.R-project.org/package=quantreg`. R package version 5.33.

O. Linton, K. Song, and Y.-J. Whang. An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154:186–202, 2010.

MOSEK ApS. *MOSEK Rmosek Package. Version 8.1.*, 2018. URL `https://docs.mosek.com/8.1/rmosek/index.html`.