Sankhyā : The Indian Journal of Statistics
Special Issue on Quantile Regression and Related Methods
2005, Volume 67, Part 2, pp 253-276

Subsampling Inference on Quantile Regression Processes

Victor Chernozhukov and Iván Fernández-Val
Massachusetts Institute of Technology, Cambridge, USA

Abstract

In program evaluation studies, important hypotheses concerning how a treatment or a social program affects the distribution of an outcome of interest can be tested using statistics derived from empirical conditional quantile processes. This paper develops simple and practical tests for verifying these hypotheses. The critical values for these tests are obtained by subsampling appropriately recentered empirical quantile regression processes. The resulting tests have not only good power and size properties, but also a much wider applicability than the available methods based on Khmaladzation. Of independent interest is also the use of recentering in subsampling, which leads to substantial improvements in the finite-sample power of the tests relative to the canonical (uncentered) subsampling. This can be attributed theoretically to an improvement in Bahadur efficiency that the recentering provides in the testing context. The new inference approach is illustrated through a reanalysis of the Pennsylvania reemployment bonus experiment.

AMS (2000) *subject classification.*
Keywords and phrases.

1 Introduction

Beginning with the early work of Quetelet on growth charts, conditional quantile models have provided a valuable method of statistical analysis. This is especially true for program evaluation studies in biometrics and econometrics, where conditional quantile methods help analyse how treatments or social programs affect the outcome distributions of interest. Basic approaches for conditional quantile estimation, proposed in Bhattacharya (1963), Chaudhuri (1991), Doksum (1974), Koenker and Bassett (1978), and Portnoy (1997) facilitate this analysis. In addition, there are methods, developed in Buchinsky and Hahn (1998), Honoré, Khan and Powell (2002), Portnoy (2002) and Powell (1986), that account for censoring of the outcome variable as well as methods, suggested in Abadie, Angrist, and Imbens

(2002), Chernozhukov and Hansen (2005), Chesher (2003), and Firpo (2004), that deal with other types of sample selectivity and endogeneity.

This paper develops inferential methods based on conditional quantile estimators, with special emphasis on the tests of hypotheses that arise in program evaluation studies. In particular, we focus on (i) tests of a stochastic dominance, (ii) tests of treatment effect significance, (iii) tests of treatment effect heterogeneity, and (iv) other specification tests (see, e.g., Abadie, 2002, Heckman, Smith and Clements, 1997, and McFadden, 1989 for additional motivation). These hypotheses involve conditional quantile functions and other parameters of the conditional distribution. The tests are formulated like Kolmogorov-Smirnov (KS) and Cramér-von-Misses-Smirnov (CMS) type tests, based on empirical conditional quantile functions. However, in many interesting applications the presence of estimated nuisance parameters, dependent data, or other features of the model jeopardizes the "distribution-free" character of these tests (this problem is usually referred to as the Durbin problem), which makes it difficult to use these tests in empirical work.

In the case of estimated nuisance parameters, the Durbin problem can be overcome by using a martingale transformation proposed by Khmaladze (1988). This procedure uses recursive projections to annihilate the component in the inference process due to the estimation of the unknown nuisance parameters, yielding a martingale that has a standard limit distribution. Koenker and Xiao (2002) have developed the Khmaladzation procedure for conditional quantiles.

Here we suggest a simple resampling alternative that (i) does not require the somewhat complex Khmaladze transformation, (ii) does not require the estimation of the nonparametric nuisance functions needed in Khmaladze approach, (iii) applies beyond the (local-to) location-scale regression models required in Koenker and Xiao (2002), (iv) has good size and better power than Khmaladzation, at least in the examples considered, (v) is robust to general forms of serial dependence in the data, to which Khmaladzation is not immune, and (vi) is computationally and practically attractive. In summary, the approach is a useful complement to Khmaladzation and is aimed at substantively expanding the scope of empirical inference using quantile regression methods.

Our testing approach is based on KS and CMS statistics, which are defined on the empirical quantile regression process. The critical values are obtained by subsampling the "mimicking" KS and CMS statistics, which are defined on an appropriately recentered empirical quantile regression process. Suppose that the original KS or CMS statistics have a limit distribution H

under the null hypothesis. Then, the mimicking statistics have distribution H under both the null and local alternatives. Hence, in order to estimate H correctly, regardless of whether the null is true or fails locally, we resample the mimicking statistics. As a result, H as well as the entire null law of the empirical quantile process are correctly estimated under local departures from the null. It should be noted that our procedure differs from the canonical subsampling for testing procedures, developed in Politis and Romano (1994), precisely in the use of recentering inside the resampled statistics. This recentering not only improves considerably the finite sample power of the subsampling based tests, but also makes the performance of the method practically insensitive to the choice of the subsample size.

To illustrate the utility of the approach, we briefly revisit the Pennsylvania re-employment bonus experiment conducted in the 1980's by the U.S. Department of Labor in order to test the incentive effects of alternative compensation schemes for unemployment insurance (UI). In these controlled experiments, UI claimants were randomly offered a cash bonus if they found a job within some prespecified period of time. Our goal is to evaluate the impact of such a scheme on the distribution of the unemployment duration. We find that the bonus offer creates a first order stochastic dominance effect on this distribution, reducing durations at nearly all quantiles. This result usefully complements other findings reported in Koenker and Xiao (2002).

The rest of the paper is organized as follows. Section 2 introduces the general testing problem, gives some illustrative examples, and describes the Durbin problem in the context of quantile process inference. Section 3 presents the subsampling test and its implementation. Sections 4 and 5 give numerical examples and an empirical application.

2 The Testing Problem

Important questions posed in the econometric and statistical literature concern the nature of the impact of a policy intervention or treatment on the outcome distributions of interest; such as, for example, whether a policy exerts a significant effect, a constant vs. heterogeneous effect, or a stochastically dominant effect; see, e.g., Doksum (1974), Koenker and Xiao (2002), Abadie (2002), Heckman, Smith and Clements (1997), and McFadden (1989). Methods based on conditional quantiles offer a good way of learning about these distributional phenomena.

Suppose Y is a real outcome variable, and X a vector of regressors. The vector X will typically include policy variables, D, other controls, and interactions. For the sake of concreteness and to present examples below, we can think of D as just a policy or treatment indicator, and the other components of X, denoted by X_{-1}, as a set of other observed characteristics. The testing framework discussed below can be easily adjusted to incorporate more general models and interactions.

Let $F_{Y|X}(y)$ and $F^{-1}_{Y|X}(\tau) = \inf\{y : F_{Y|X}(y) \geq \tau\}$ denote the conditional distribution function and the τ-quantile of Y given X, respectively. The basic conditional quantile model, introduced in Hogg (1975), takes the linear form:

$$F^{-1}_{Y|X}(\tau) = X'\beta_n(\tau),$$

for all quantiles of interest $\tau \in \mathcal{T}$, where $\mathcal{T}$ is a closed subinterval of $(0, 1)$. In order to facilitate a local power analysis, the parameter $\beta_n(\tau)$ is allowed to depend on the sample size, n. This specification corresponds to a random coefficients model $Y = X'\beta_n(U)$, where $U \sim U(0, 1)$ conditional on X. This is a standard model in quantile regression analysis and allows the regressors to affect the entire shape of the conditional distribution, encompassing the classical (location-shift) regression model as a special case.

We consider the following general null hypothesis:

$$R(\tau)\beta_n(\tau) - r(\tau) = 0, \quad \tau \in \mathcal{T}, \tag{1}$$

where $R(\tau)$ denotes $q \times p$ matrix, with $q \leq p = \dim(\beta)$, and $r(\tau)$ is a $q \times 1$ vector. We assume that the functions $R(\tau)$ and $r(\tau)$ are continuous in τ over $\mathcal{T}$, and that $\beta_0(\tau) \equiv \lim_n \beta_n(\tau)$ exists and is continuous in τ over $\mathcal{T}$. In the examples of interest, the components $R(\tau)$ and $r(\tau)$ are defined as functions of the conditional distribution and thus need to be estimated. The estimates will be denoted as $\widehat{R}(\tau)$ and $\widehat{r}(\tau)$. This hypothesis embeds several interesting hypotheses about the parameters of the conditional quantile function, as illustrated in the examples presented below.

Our discussion focuses on tests derived from the quantile regression process $\widehat{\beta}_n(\cdot)$ defined, following Koenker and Bassett (1978), as

$$\widehat{\beta}_n(\tau) = \arg\min_{\beta \in \mathbb{R}^d} \sum_{t=1}^{n} \rho_\tau\left(Y_i - X_i\beta\right), \quad \tau \in \mathcal{T}, \tag{2}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$. Other estimators mentioned in the introduction could also be considered instead, depending on the problem at hand. The regularity conditions presented below allow for a rich variety of underlying conditional quantile estimators.

We next consider the basic inference (empirical) process:

$$v_n(\tau) = \widehat{R}(\tau)\widehat{\beta}_n(\tau) - \widehat{r}(\tau) - \Psi(\tau), \tag{3}$$

and derive from it test statistics, $S_n = f(v_n(\cdot))$, where

$$S_n = \sqrt{n} \sup_{\tau \in \mathcal{T}} \|v_n(\tau)\|_{\widehat{V}(\tau)}, \quad S_n = n \int_{\mathcal{T}} \|v_n(\tau)\|^2_{\widehat{V}(\tau)} d\tau; \tag{4}$$

for KS and CMS, respectively. Here, $\|a\|_V$ denotes $\sqrt{a'Va}$; $\widehat{V}(\tau)$ is a symmetric weight matrix, such that $\widehat{V}(\tau) = V(\tau) + o_p(1)$, uniformly in τ, where $V(\tau)$ is a positive definite and continuous symmetric matrix, uniformly in τ. Section 3 provides details concerning the choice of $V(\tau)$ and $\widehat{V}(\tau)$.

In order to introduce some specific examples, we consider the model $F^{-1}_{Y|D,X}(\tau) = D\beta_1(\tau) + X'_{-1}\beta_{-1}(\tau)$, where D denotes a policy or treatment variable, the impact of which we would like to analyse. In particular, D can be thought of as an indicator of treatment receipt or of program participation.

Example 1. (*The Hypothesis of a Significant Effect.*) A basic hypothesis is that the treatment impact summarized by $\beta_1(\tau)$ differs from zero at least for some τ. This hypothesis fits in the stated general hypothesis with $R(\tau) = R = [1, 0, ...]$ and $r(\tau) = 0$.

Example 2. (*The Hypothesis of a Constant Effect vs. Heterogeneous Effects.*) Another important hypothesis is whether the treatment impact does not vary across quantiles, i.e. $\beta_1(\tau) = \beta$ for some unknown β for all τ. In this case, $R(\tau) = R = [1, 0, ...]$ and $r(\tau) = \beta$. We can estimate the component $r(\tau) = \beta$ in this case by $\int_{\mathcal{T}} \widehat{\beta}_{1n}(\tau)$. The alternative is the hypothesis of heterogeneous effect, i.e. $\beta_1(\tau)$ varies across τ.

Example 3. (*The Hypothesis of Stochastic Dominance*). The test of *stochastic dominance* in the model stated above involves the dominance composite null $\beta_1(\tau) \geq 0$, for all $\tau \in \mathcal{T}$, versus the non-dominance alternative $\beta_1(\tau) < 0$, for some $\tau \in \mathcal{T}$. In this case, the least favourable null involves $r(\tau) = 0$ and $R = -[1, 0...]$. A signed version of the KS statistic,

$$S_n = \sqrt{n} \cdot \sup_{\tau \in \mathcal{T}} \left(-\text{sign } (\widehat{\beta}_{1n}(\tau)) \|\widehat{\beta}_{1n}(\tau)\|_{\widehat{V}(\tau)} \right),$$

can be used to test this hypothesis.

In what follows, we use P_n^* to denote outer probability, which possibly depends on the sample size n, and $\Rightarrow$ to denote weak convergence under P_n^* in the space of bounded functions $\ell^\infty(\mathcal{T})$. We maintain the following main assumptions:

A.1 For each n, $(Y_t, X_t, t \geq 1)$ is a stationary and strongly mixing sequence on the probability space $(\Omega, \mathcal{F}, P_n)$.[1]

A.2 The law of the data $(Y_t, X_t, 1 \leq t \leq n)$, denoted by $P_n^{[n]}$, is contiguous to $P^{[n]}$, for some fixed probability measure P. Furthermore, $R(\tau)\beta_n(\tau) - r(\tau) = \Psi(\tau) + g(\tau)$, where for a fixed continuous function $p(\tau) : \mathcal{T} \to \mathbb{R}^q$ either (a) $g(\tau) = p(\tau)/\sqrt{n}$ for each n, or (b) $g(\tau) = p(\tau) \not\equiv 0$ for each n.

A.3 (a) Under any local alternative A.2(a), the quantile estimates and nuisance parameter estimates satisfy: $\sqrt{n}(\widehat{\beta}_n(\cdot) - \beta_n(\cdot)) \Rightarrow b(\cdot)$, $\sqrt{n}(\widehat{R}(\cdot) - R(\cdot)) \Rightarrow \rho(\cdot)$, $\sqrt{n}(\widehat{r}(\cdot) - r(\cdot)) \Rightarrow \varsigma(\cdot)$, jointly in $\ell^\infty(\mathcal{T})$, where (b, ρ, ς) is a zero mean continuous Gaussian process with a non-degenerate covariance kernel. (b) Under any global alternative A.2(b), the same holds, except that the limit $(\tilde{b}, \tilde{\rho}, \tilde{\varsigma})$ does not need to have the same distribution as in A.3(a), and may depend on the alternative.

These assumptions immediately yield the following proposition, which describes the limit distribution of the inference process and derived tests statistics.

Proposition 1. *1. Under conditions A.1, A.2(a), A.3, in $\ell^\infty(\mathcal{T})$,*

$$\sqrt{n}v_n(\cdot) \Rightarrow v(\cdot) = v_0(\cdot) + p(\cdot), \quad v_0(\cdot) \equiv u(\cdot) + d(\cdot),$$

where $u(\tau) = R(\tau)'b(\tau)$ and $d(\tau) = (\beta_0(\tau)\rho(\tau) - \varsigma(\tau))$. Under the null hypothesis, $p = 0$, the test statistic S_n, defined in (4), converges in distribution to $S \equiv f(v_0(\cdot))$.

2. Under the conditions A.1, A.2(b), A.3, $\sqrt{n}(v_n(\cdot) - g(\cdot)) \Rightarrow \tilde{v}(\cdot) \equiv \tilde{u}(\cdot) + \tilde{d}(\cdot)$, where $\tilde{u}(\tau) = R(\tau)'\tilde{b}(\tau)$ and $\tilde{d}(\tau) = (\tilde{\beta}_0(\tau)\tilde{\rho}(\tau) - \tilde{\varsigma}(\tau))$. Moreover, S_n converges in probability to $+\infty$.

Let us first discuss the conditions required for this proposition. Condition A.1 allows for a wide variety of data processes, including i.i.d. and stationary time series data. Strong mixing is sufficient, but not necessary, for consistency of subsampling. Stationarity can be replaced by more general stability conditions, see e.g. Chapter 4 in Politis, Romano and Wolf (1999).

Conditions A.2(a) and A.2(b) formulate local and global alternatives. Condition A.3 is a general condition that requires the parameters of the

[1] Recall that a stationary series $(Y_t, X_t, t \geq 1)$ is strongly mixing if $\sup_{A,B,m\geq 1} |P_n(A \cap B) - P_n(A)P_n(B)| \to 0$ as $k \to \infty$, for A and B ranging over σ-fields generated respectively by $(Y_t, X_t, 1 \leq t \leq m)$ and $(Y_t, X_t, m + k \leq t < \infty)$.

empirical conditional quantile processes and of the null hypothesis to be asymptotically Gaussian processes. These conditions are implied by a variety of conditions in the literature; see, e.g., Portnoy (1991) for the treatment of quantile regression processes (2) under general forms of dependence and heterogeneity.

Assumptions A.1-A.3 are substantially more general than the assumptions imposed in Koenker and Xiao (2002) in order to implement the Khmaladzation approach, which include i.i.d. sampling and (local-to) a location-scale model in all the covariates. These stronger assumptions are not necessary in any of the previous examples. The Khmaladzation method developed in Koenker and Xiao (2002) does not apply outside these special settings.

Proposition 1 shows that the limit inference process $v(\cdot)$ is the sum of three components, $u(\cdot)$, $d(\cdot)$, and $p(\cdot)$. The *usual* component $u(\cdot)$ is typically a Gaussian process with a non-standard covariance kernel, so its distribution can not be feasibly simulated. This problem may be assumed away by imposing i.i.d. conditions. However, the problem does not go away, once the data is a general time series or the quantile regression model is misspecified (in which case the model is interpreted as an approximation). Portnoy (1991) provides expressions for the covariance kernel in time series settings, and Angrist, Chernozhukov and Fernández-Val (2005) analyse the misspecified case. In such important setting, the Khmaladzation approach is not valid in its present form. The component $d(\cdot)$ is the *Durbin* component that is present because $R(\cdot)$ and $r(\cdot)$ are estimated. Khmaladzation is often used to eliminate this component. The component $p(\cdot)$, which describes deviations from the null, determines the power of the tests. As Koenker and Xiao (2002) show, the Khmaladzation inadvertently removes some portion of this component.

Khmaladzation requires estimation of several nonparametric nuisance functions that appear as deterministic components of $d(\cdot)$. Koenker and Xiao (2002) develop their procedure for a location-scale shift model, where the nuisance function is scalar-valued by assumption, greatly facilitating the implementation.

In the next section, we describe a simple approach that is very useful in practice and does not require estimation of nuisance functions. The approach has a much wider applicability than Khmaladzation does.

3 Resampling Test and Its Implementation

3.1. The test. Our approach is based on the "mimicking" inference process $\bar{v}(\cdot)$ and the corresponding test statistic $\mathcal{S}_n$:

$$\bar{v}_n(\tau) = v_n(\tau) - g(\tau), \quad \mathcal{S}_n = f(\bar{v}_n(\cdot)).$$

PROPOSITION 2. *Given A.1, A.2(a), A.3* $\sqrt{n}\bar{v}_n(\cdot) \Rightarrow v_0(\cdot)$, *and* $\mathcal{S}_n \Rightarrow S$. *2. Given A.1, A.2(b), A.3* $\sqrt{n}\bar{v}_n(\tau) \Rightarrow \bar{v}(\cdot) = \tilde{u}(\cdot) + \tilde{d}(\cdot)$, *and* $\mathcal{S}_n \Rightarrow \mathcal{S} \equiv f(\bar{v}(\cdot))$.

Under local alternatives, the statistic $\mathcal{S}_n$ correctly mimics the null behaviour of S_n, even when the null hypothesis is false. This does not happen under global alternatives, but this is not important for the consistency of the test. In what follows we use $v_n(\tau)$ itself to "estimate" $g(\tau)$, and use subsampling to estimate consistently the distribution of $\mathcal{S}$, which equals that of S under the null hypothesis. We describe the test in two steps.

STEP 1. For cases when $W_t = (Y_t, X_t)$ is i.i.d., construct all possible subsets of size b. The number of such subsets B_n is "n choose b." For cases when $\{W_t\}$ is a time series, construct $B_n = n - b + 1$ subsets of size b of the form $\{W_i, ..., W_{i+b-1}\}$. Compute the inference process $v_{n,b,i}(\cdot)$, for each i-th subset, $i \leq B_n$. (In practice, a smaller number B_n of randomly chosen subsets can also be used, provided that $B_n \to \infty$ as $n \to \infty$).

Denote by v_n the inference process computed over the entire sample; and by $v_{n,b,i}$ the inference process computed over the i-th subset of data, and define:

$$\widehat{\mathcal{S}}_{n,b,i} \equiv \sup_{\tau \in \mathcal{T}} \sqrt{b} \| v_{n,b,i}(\tau) - v_n(\tau) \|_{\widehat{V}(\tau)}$$

or,

$$\widehat{\mathcal{S}}_{n,b,i} \equiv b \int_{\mathcal{T}} \| v_{n,b,i}(\tau) - v_n(\tau) \|^2_{\widehat{V}(\tau)} d\tau,$$

for cases when S_n is KS or CMS statistics, respectively. Define

$$G(x) \equiv Pr\{\mathcal{S} \leq x\} \quad \text{and} \quad H(x) \equiv Pr\{S \leq x\}.$$

Note that as $b/n \to 0$ and $b \to \infty$, $\sqrt{b}\|v_n(\cdot) - g(\cdot)\| = \sqrt{b} \cdot O_p(1/\sqrt{n}) = o_p(1)$, including when $g(\tau) = p(\tau)/\sqrt{n}$. As a result, $\sqrt{b}\|(v_{n,b,i}(\cdot) - g(\cdot)) + (g(\cdot) - v_n(\cdot))\| = \sqrt{b}\|v_{n,b,i}(\cdot) - g(\cdot)\| + o_p(1)$, uniformly in i. The distribution of $\widehat{\mathcal{S}}_{n,b,i}$ can then consistently estimate G, which coincides with H under local alternatives. Thus, the following step is clear.

STEP 2. Estimate $G(x)$ by $\widehat{G}_{n,b}(x) = B_n^{-1}\sum_{i=1}^{B_n} 1\{\widehat{\mathcal{S}}_{n,b,i}(\tau) \leq x\}$. Obtain the critical value as the $(1-\alpha)$-th quantile of $\widehat{G}_{n,b}$, $c_{n,b}(1-\alpha) = \widehat{G}_{n,b}^{-1}(1-\alpha)$. Finally, reject the null hypothesis if $S_n > c_{n,b}(1-\alpha)$.

THEOREM 1. *Given A.1 - A.3 as* $b/n \to 0, b \to \infty, n \to \infty$, $B_n \to \infty$,

(i) When the null hypothesis is true, $p = 0$, *if* H *is continuous at* $H^{-1}(1-\alpha)$*:*

$$c_{n,b}(1-\alpha) \xrightarrow{P_n^*} H^{-1}(1-\alpha), \quad P_n^*(S_n > c_{n,b}(1-\alpha)) \to \alpha.$$

(ii) Under any local alternative A.2(a), $p \not\equiv 0$, *if* H *is continuous at* $H^{-1}(1-\alpha)$*:*

$$c_{n,b}(1-\alpha) \xrightarrow{P_n^*} H^{-1}(1-\alpha), \quad P_n^*(S_n \leq c_{n,b}(1-\alpha)) \to 1-\beta,$$

where $\beta = Pr\left(f(v_0(\cdot)+p(\cdot)) > H^{-1}(1-\alpha)\right) > \alpha$.

(iii) Under the global alternative A.2(b), if G *is continuous at* $G^{-1}(1-\alpha)$*:*

$$c_{n,b}(1-\alpha) \xrightarrow{P_n^*} G^{-1}(1-\alpha), \quad P_n^*(S_n \leq c_{n,b}(1-\alpha)) \to 0.$$

(iv) $H(x)$ *and* $G(x)$ *are continuous if the covariance function of* $v(\cdot)$ *and* $\bar{v}(\cdot)$ *is nondegenerate.*

Theorem 1 shows that the test based on subsampling asymptotically has correct level, is consistent against global alternatives, has nontrivial power against root-n alternatives, and has the same power as the test with known critical value. Furthermore as $\|p(\tau)\| \to \infty$, the power β goes to one.

It is useful to comment on the use of recentering in subsampling. Under global alternatives, the critical value $c_{n,b}$ used in the subsampling test with recentering converges to a constant; in contrast, the critical value obtained by the canonical subsampling test without recentering (Politis, Romano and Wolf, 1999 diverges to ∞ at the rate $\sqrt{b}$. Therefore, both tests are consistent, but this observation suggests that the canonical test should be less efficient in the Bahadur sense and, of course, less powerful in finite samples. Computational experiments confirm this observation. One should note that, under local alternatives, the critical values of both tests converge to the same value. Therefore, both tests have identical Pitman efficiency.

3.2. Practical considerations. It may be sometimes more practical to use a grid $\mathcal{T}_n$ in place of $\mathcal{T}$ with the largest cell size $\delta_n \to 0$ as $n \to \infty$.

Corollary 1. *Propositions 1 and 2 and Theorems 1 and 2 are valid for the piece-wise constant approximations of the finite-sample processes, given that $\delta_n \to 0$ as $n \to \infty$.*

3.3. Estimation of $V(\tau)$. In order to increase the testing power we could set

$$V(\tau) = [\Omega(\tau)]^{-1} \equiv \mathrm{Var}\,[\bar{v}(\tau)]^{-1},$$

which is an Anderson-Darling type weight. It is also convenient to use subsampling itself to estimate $\Omega(\tau)$. Under asymptotic integrability conditions, the subsampling variance matrix is a consistent estimate of $\Omega(\tau)$. Without integrability conditions, we can estimate $\Omega(\tau)$ using trimmed moments in conjunction with asymptotic normality. Another useful method is the quantile approach that uses interquantile ranges of the components $\bar{v}_j(\tau)$ and $\bar{v}_k(\tau) + \bar{v}_j(\tau)$, along with normality, to determine the variance matrix of $\bar{v}(\tau) = (\bar{v}_j(\tau), j = 1, ..., p)$. We shall focus only on the first approach for the sake of brevity. This is the approach we use in simulations and in the empirical section. However, the validity of the trimmed moments and quantile approaches is immediate from Theorem 2 that shows the uniform in τ consistency of the estimates of the key ingredients.

Theorem 2 establishes four results. The first result states that the subsampling law of $\sqrt{b}\,(v_{i,b,n}(\cdot) - v_n(\cdot))$ converges weakly to that of $\bar{v}(\cdot)$, which is a direct consequence of the ingenious results of Politis, Romano and Wolf (1999). The second result provides consistent estimates of the truncated moments of $\bar{v}(\cdot)$, using the corresponding quantities of the subsampling law. The third result shows the convergence of untrimmed moments, assuming uniform integrability conditions. The fourth states the convergence of subsampling percentiles uniformly in the quantile index τ.

Define the notation $x^m \equiv (x_1, ..., x_p)^m \equiv x_1^{m_1} \times ... \times x_p^{m_p}$ for any $m = (m_1, ..., m_p)$ valued on nonnegative integers. A simple estimate of the (untrimmed) moment is given by:

$$E_{L_{n,b}} \left[\sqrt{b}\,(v_{i,b,n}(\tau) - v_n(\tau))\right]^m = \frac{1}{B_n} \sum_{i=1}^{B_n} \left[\sqrt{b}\,(v_{i,b,n}(\tau) - v_n(\tau))\right]^m,$$

where $L_{n,b}$ denotes the sub-sampling law of $\sqrt{b}(v_{i,b,n}(\cdot) - v_n(\cdot))$. Using various m we can obtain estimates of variance, covariance, and other moments of $\bar{v}(\tau)$.

Next, in order to discuss the trimmed moments as well as forthcoming formal results, define the following notation. Let $\tau \mapsto v(\tau)$ be an element of

$\ell^\infty(\mathcal{T})$, and $\mathbf{L}(c,k)$ be a class of Lipschitz functions $\varphi : \ell^\infty(\mathcal{T}) \to \mathbb{R}^K$ that satisfy:

$$\|\varphi(v) - \varphi(v')\| \leq c \cdot \sup_{\tau} \|v(\tau) - v'(\tau)\|, \quad \|\varphi(v)\| \leq k,$$

where c and k are suitably chosen positive constants. For probability laws Q and Q', define the bounded Lipschitz metric as $\rho_{BL}(Q,Q') = \sup_{\varphi \in \mathbf{L}} \|E_Q\varphi - E_{Q'}\varphi\|$.

Consider a trimming function $f_K(x) \equiv \min(\max(-K,x),K)$, where $K > 0$ and $x \in \mathbb{R}$. A motivating example for φ is $\varphi(v) = f_{K(\tau)}(v(\tau)^m)$, where $K(\tau) \leq K$ for all τ, which defines all kinds of trimmed moments and correlations. E.g.

$$E_{L_{n,b}}\left[f_K\left\{\left[\sqrt{b}\,(v_{i,b,n}(\tau) - v_n(\tau))\right]^m\right\}\right] = \frac{1}{B_n}\sum_{i=1}^{B_n}\left[f_K\left\{\left[\sqrt{b}\,(v_{i,b,n}(\tau) - v_n(\tau))\right]^m\right\}\right],$$

where $L_{n,b}$ symbolically denotes the subsampling law of $\sqrt{b}(v_{i,b,n}(\cdot) - v_n(\cdot))$.

THEOREM 2. *Under conditions A.1-A.3*

(i) letting L and L_0 denote the laws of $\bar{v}(\cdot)$ and $v_0(\cdot)$ in $\ell^\infty(\mathcal{T})$, respectively,

$$\rho_{BL}(L_{n,b}, L) \xrightarrow{P_n^*} 0,$$

and L equals L_0 under local alternatives.

(ii)

$$\sup_{\tau\in\mathcal{T}}\left|E_{L_{n,b}}\left[f_K\left\{\left[\sqrt{b}(v_{i,b,n}(\tau) - v_n(\tau))\right]^m\right\}\right] - E_L\left[f_K\left\{\left[\bar{v}(\tau)\right]^m\right\}\right]\right| \xrightarrow{P_n^*} 0.$$

(iii) The trimming in (ii) can be dropped if uniformly in $P \in \{P_n, n > n_0\}$ for some constants n_0 and n_1:[2]

$$\left\{\sup_{\tau\in\mathcal{T}} |\{\sqrt{J}\,(v_J(\tau) - g_P(\tau))\}^m|, J \geq n_1\right\} \quad \textit{are uniformly integrable.}$$

[2] In this formula the departure from the null in A.2 $g(\tau)$ is indexed by probability measure P to emphasize this dependence for clarity.

(iv) For $\bar{\nu}(\tau) \equiv \sqrt{b}\,(v_{i,n,b}(\tau) - v_n(\tau))$ *for all* j, k

$$\sup_{\tau\in\mathcal{T}} \left|Q_{\bar{\nu}(\tau)_j}(p|L_{b,n}) - Q_{\bar{v}(\tau)_j}(p|L)\right| \xrightarrow{P_n^*} 0,$$
$$\sup_{\tau\in\mathcal{T}} \left|Q_{\bar{\nu}(\tau)_j+\bar{\nu}(\tau)_k}(p|L_{b,n}) - Q_{\bar{v}(\tau)_j+\bar{v}(\tau)_k}(p|L)\right| \xrightarrow{P_n^*} 0,$$

if $v(\cdot)$ *has a nondegenerate covariance function. Here* $Q_Z(p|L)$ *denotes the* p*-th quantile of the random scalar* Z *under the probability law* L*.*

The integrability imposed in (iii) is a fairly weak condition in practical situations. Under stronger uniform integrability conditions than in (iii), imposed on the higher order powers of the inference process, it is possible to strengthen the convergence in (iii) to convergence in the mean squared sense. This can be done by adopting the arguments of Carlstein (1986) and Fukuchi (1999) developed for the random variable case.

3.4. Choice of block size. In Sakov and Bickel (2000) and in Politis, Romano and Wolf (1999) various rules are suggested for choosing the appropriate subsample size. Politis, Romano and Wolf (1999) focus on the calibration and minimum volatility methods. The calibration method involves picking the optimal block size and appropriate critical values on the basis of simulation experiments conducted with a model that approximates a situation at hand. The minimum volatility method involves picking (or combining) among the block sizes that yield more stable critical values. More detailed suggestions emerge from Sakov and Bickel (2000), who suggest choosing $b \propto n^{1/2}$ in the case of the replacement version of subsampling applied to a sample quantile. Our own experiments indicated that this rule works well in the present context. Specifically, we used a similar rule $b = m + n^{1/c}$, where m is the minimal plausible sample size and c can be set to $1/2$. In the simulation experiments, we also use $c = 1/4$ to have cheaper computations.

4 A Computational Example

In this section we conduct computational experiments to assess the size and power properties of the tests in finite samples. To help us compare the performance of the resampling tests with the alternative tests based on Khmaladzation without prejudicing against the latter, we use the same design as in Koenker and Xiao (2002). In particular, we consider 36 different scenarios of the location-shift hypothesis in Example 2. The data are generated from the model:

$$\begin{aligned}
&Y_i = \alpha + \beta D_i + \sigma(D_i) \cdot \epsilon_i,\\
&\sigma(D_i) = \gamma_0 + \gamma_1 \cdot D_i,\\
&\epsilon_i \sim N(0,1),\ D_i \sim N(0,1),\\
&\alpha = 0, \beta = 1, \gamma_0 = 1.
\end{aligned}$$

Under the null hypothesis $\gamma_1 = 0$. We examine the empirical rejection frequencies for 5% nominal level tests for different choices of sample size n, sub-sample block b, and heteroscedasticity parameter γ_1. The sample size ranges from $n = 100$ to $n = 800$. To demonstrate the robustness of the technique with respect to the sub-sample size, we use the replacement version of subsampling with choices of the block size $b = 20 + n^{1/4}$, $b = 20 + n^{1/(2+\epsilon)}$,[3] for $\epsilon = 0.01$, $b = n/4$, and $b = n$; where $b = n$ corresponds to the n out of n bootstrap. In constructing the test, we use the OLS estimate for $\widehat{\beta}$ and an equally spaced grid of 19 quantiles $\mathcal{T}_n = \{.05, .10, ..., .95\}$. When $\gamma_1 = 0$ the model is a location-shift model, and the rejection rates yield the empirical sizes. When $\gamma_1 \neq 0$ the model is heteroscedastic, and the rejection rates give the empirical powers. We use 1000 replications in each of the simulations, and employ $B_n = 250$ bootstrap repetitions within each replication to compute the critical values.[4]

TABLE 1. EMPIRICAL REJECTION FREQUENCIES FOR 5% LEVEL KHMALADZE TEST (KOLMOGOROV-SMIRNOV STATISTIC)*

	H = .5 Bofinger			H = .6 Bofinger		
	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$
$n = 100$	.101	.264	.804	.035	.211	.755
$n = 400$	.054	.812	1	.043	.809	1
$n = 800$	.053	.982	1	.050	.969	1
	H = .7 Bofinger			H = 1 Bofinger		
	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$
$n = 100$	.016	.126	.641	.009	.053	.197
$n = 400$	.035	.632	1	.023	.412	.997
$n = 800$	.049	.924	1	.041	.792	1

*Results extracted from Koenker and Xiao (2002). H denotes the different bandwidth choices relative to Bofinger rule.

[3]The replacement and nonreplacement subsampling here are equivalent with probability converging to 1 if $b^2/n \to 1$, so our formal results cover only these first two schemes.

[4]The maximal simulation standard error for the empirical sizes and powers of the tests is $\max_{0 \le p \le 1} \sqrt{p(1-p)/1000} \approx .016$.

Table 1 reports the results for the KS test based on Khmaladzation, extracted from Koenker and Xiao (2002), for the different values of n and γ_1 considered here. Tables 2 and 3 give the results for the subsampling tests based on CMS statistic and KS statistic, respectively. For the CMS statistic we add an adjustment of 1.96 times an estimate of the standard error of the subsampling quantile estimate $c_{n,b}$ to account for the finite number of bootstrap repetitions B_n. For the KS statistic, we find that the bias has the opposite direction. We adjust the quantile estimate $c_{n,b}$ by subtracting 1.96 times the standard error. For each statistic, we calculate rejection frequencies using the canonical (uncentered) subsampling test of Politis, Romano, and Wolf (1999) (Panel B), and the recentered version proposed here (Panel A).

Table 2. Empirical Rejection Frequencies for 5% resampling test (Cramér-Von-Mises-Smirnov Statistic), using 250 bootstrap draws and 1000 repetitions.*

	$b = 20 + n^{1/4}$			$b = 20 + n^{1/2.01}$		
	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$
	A - CENTERED					
$n = 100$	.012	.454	.996	.038	.497	.996
$n = 400$	.049	.991	1	.057	.996	1
$n = 800$	.059	1	1	.062	1	1
	B - UNCENTERED (CANONICAL)					
$n = 100$	.009	.185	.888	.015	.166	.855
$n = 400$	.043	.986	1	.043	.987	
$n = 800$	.057	1	1	.053	1	1

	$b = n$			$b = n/4$		
	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$
	A - CENTERED					
$n = 100$	.030	.502	.998	.029	.483	.994
$n = 400$	.062	.997	1	.063	.995	1
$n = 800$	.076	1	1	.078	1	1
	B - UNCENTERED (CANONICAL)					
$n = 100$	0	0	0	.013	.212	.889
$n = 400$	0	0	.159	.011	.948	1
$n = 800$	0	.001	.768	.011	1	1

*All the results are reproducible and the programs are available from the authors. Maximal simulation s.e. = .016.

Overall, the resampling tests are quite powerful and have small size distortions even in small samples. The results also suggest that resampling procedures substantively outperform the Khmaladzation procedures in terms of power. Notably, for the model considered, even using a very small subsample size, $b = 20 + n^{1/4}$, leads to reliable, powerful, and computationally attractive inference. Our main proposal, the recentered subsampling test, in addition to having a considerably better power than other methods, makes the resampling method quite robust to variations of subsample size. This is not the case for the uncentered subsampling.

TABLE 3. EMPIRICAL REJECTION FREQUENCIES FOR 5% RESAMPLING TEST (KOLMOGOROV-SMIRNOV STATISTIC), USING 250 BOOTSTRAP DRAWS AND 1000 REPETITIONS.*

	$b = 20 + n^{1/4}$			$b = 20 + n^{1/2.01}$		
	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$
			A - CENTERED			
$n = 100$	.005	.217	.955	.009	.221	.962
$n = 400$	.026	.962	1	.023	.965	1
$n = 800$	.032	1	1	.046	1	1
			B - UNCENTERED (CANONICAL)			
$n = 100$	.004	.116	.684	.007	.084	.572
$n = 400$	.038	.947	1	.021	.923	1
$n = 800$	.032	1	1	.049	1	1

	$b = 20 + n^{1/4}$			$b = 20 + n^{1/2.01}$		
	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$	$\gamma = 0$	$\gamma = .2$	$\gamma = .5$
			A - CENTERED			
$n = 100$	.005	.207	.975	.008	.213	.960
$n = 400$	.020	.976	1	.028	.979	1
$n = 800$	.036	1	1	.041	1	1
			B - UNCENTERED (CANONICAL)			
$n = 100$	0	0	0	.006	.108	.675
$n = 400$	0	0	0	.010	.807	1
$n = 800$	0	0	0	.016	.994	1

*All the results are reproducible and the programs are available from the authors. Maximal simulation s.e. = .016.

Table 4 reports empirical sizes for the resampling tests using larger sample sizes. Here, the smallest sub-sample blocks yield good sizes, whereas increasing the sub-sample blocks can create size distortions of up to 4% in the CMS test, which also vanish quite slowly as $n \to \infty$. A potential drawback of our method, shared also by the tests based on Khmaladzation, is that some finite sample adjustments are needed to size-correct the critical values. In results not reported, we find that the adjustment is sensitive to the number of bootstrap repetitions B_n.[5] In practice, we recommend calibrating

[5]Results for Tables 2 to 4 using unadjusted tests are available from the authors upon

the adjustment to the situation at hand, or to use a conservative strategy based on the unadjusted KS statistic, which in our experience always yields empirical sizes smaller than the nominal levels of the test.

Table 4. Empirical Rejection Frequencies for 5% resampling test ($\gamma = 0$), using 250 bootstrap draws and 1000 repetitions.*

	$b = 20 + n^{1/4}$			$b = 20 + n^{1/2.01}$		
	b	CMS	KS	b	CMS	KS
	A - CENTERED					
$n = 1000$	26	.041	.029	52	.071	.046
$n = 2000$	27	.049	.038	64	.061	.043
$n = 5000$	29	.052	.029	90	.074	.060
$n = 10000$	30	.054	.036	118	.086	.068
	B - UNCENTERED (CANONICAL)					
$n = 1000$	26	.043	.026	52	.058	.045
$n = 2000$	27	.048	.038	64	.045	.040
$n = 5000$	29	.049	.030	90	.063	.059
$n = 10000$	30	.049	.036	118	.081	.068

*All the results are reproducible and the programs are available from the authors. Maximal simulation s.e. = .016.

5 An Empirical Application

To illustrate the present approach, we will re-analyse the Pennsylvania re-employment bonus experiment by expanding on the empirical questions considered in Koenker and Xiao (2002). This experiment was conducted in the 1980's by the U.S. Department of Labor[6] in order to test the incentive effects of an alternative compensation scheme for unemployment insurance (UI). In this experiment, UI claimants were randomly offered a cash bonus if they found a job within some pre-specified period of time and if the job was retained for a specified duration. The main goal was to evaluate the impact of such a scheme on the unemployment duration.

As in Koenker and Xiao (2002) we restrict attention to the compensation schedule that includes a lump-sum payment of six times the weekly unemployment benefit for claimants who establish the reemployment within 12 weeks (in addition to the usual weekly benefits). The definition of the unemployment spell includes one waiting week, with the maximum of uninterrupted full weekly benefits of 27.

request.

[6]There is a significant empirical literature focusing on the analysis of this and other similar experiments, see e.g. the review of Meyer (1995).

The model under consideration is a linear conditional quantile model for the logarithm of duration:

$$Q_{\log(T)}(\tau|X) = \alpha(\tau) + \delta(\tau)D + X'\beta(\tau),$$

where T is the duration of unemployment, D is the indicator of the bonus offer, and X is a set of socio-demographic characteristics (age, gender, number of dependents, location within the state, existence of recall expectations, and type of occupation). Further details are given in Koenker and Bilias (2001).

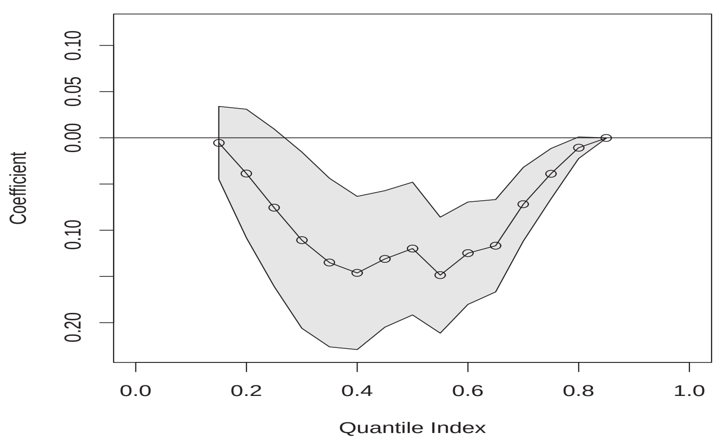

Figure 1. Quantile Treatment Effect for Unemployment Duration

The three basic hypotheses, described in Table 5, include:

- treatment effect is insignificant across a portion of the distribution ($\mathcal{T} = [.15, .85]$),
- treatment effect is constant across most of the distribution ($\mathcal{T} = [.15, .85]$),
- treatment effect is unambiguously beneficial: $\delta(\tau) < 0$ for all $\tau \in \mathcal{T}$.

We implemented the subsampling test following the procedure described in Section 3, and report the results in Table 5. We used $b = 3000$ as the subsample size and did not consider subsamples of smaller sizes, because they often yielded singular designs (many of X are dummy variables taking on positive value with probability $2 - 10\%$).[7]

[7]For this application, we use 10,000 bootstrap repetitions to obtain the critical values.

Table 5. Results of the tests for the re-employment bonus experiment, using $b = 3,000$ and recentered subsampling with replacement.*

Hypothesis	Null	Alternative	KS Statistic	5% Critical Value	Decision
No effect	$\delta(\tau) = 0$	$\delta(\tau) \neq 0$	3.76	3.21	Reject
Location shift	$\delta(\tau) = \delta$	$\delta(\tau) \neq \delta$	2.96	2.92	Reject
Dominance effect	$\delta(\tau) \leq 0$	$\delta(\tau) > 0$, for some τ	-.23	2.11	Accept

*Critical values obtained using 10,000 bootstrap repetitions.

The first two hypotheses are decisively rejected, supporting the earlier conclusions of Koenker and Xiao (2002) reached using Khmaladzation techniques. The hypothesis of stochastic dominance, the third one, is decisively supported.[8] Thus, the bonus offer creates a first order stochastic dominance effect on the unemployment duration, suggesting that the program is unambiguously beneficial from this point of view. These additional results complement, in an interesting way, the set of inference results given in Koenker and Xiao (2002).

6 Conclusion

A simple and practical resampling test is offered as an to the Khmaladzation technique, suggested in Koenker and Xiao (2002). This alternative has good power and does not require non-parametric estimation of nuisance functions. It applies to both to i.i.d. and time series data. Finite-sample experiments provide a strong evidence in support of this technique and an empirical illustration shows its utility.

Appendix A

Throughout the appendix, w.p. $\rightarrow 1$ means "with (inner) probability converging to one".

Proof of Proposition 1 and 2. The results are immediate from A.1-A.3 by the continuous mapping theorem. □

[8] Unadjusted critical values for the KS statistics lead to the same conclusions about the tests. Test based on uncentered subsampling, however, do not reject any of the hypotheses reflecting the lower power of this procedure against local departures of the null hypotheses. These results are available from the authors upon request.

PROOF OF THEOREM 1. We give the proof for the KS statistic. Extensions to other statistics defined in the text are straightforward.

I.To prove (i)- (iii), define $\dot{G}_{n,b}(x)$ and write out $\widehat{G}_{n,b}(x)$ as

$$\dot{G}_{n,b}(x) = B_n^{-1} \sum_{i \leq B_n} 1\Big[A_i \leq x\Big],$$
$$\widehat{G}_{n,b}(x) = B_n^{-1} \sum_{i \leq B_n} 1\Big[\widehat{A}_i \leq x\Big],$$

where

$$A_i = \sup_{\tau \in \mathcal{T}} \Big\| V^{1/2}(\tau) \left(\sqrt{b}(v_{i,b,n}(\tau) - g(\tau)) \right) \Big\|,$$
$$\widehat{A}_i = \sup_{\tau \in \mathcal{T}} \Big\| \widehat{V}^{1/2}(\tau) \left(\sqrt{b}(v_{i,b,n}(\tau) - g(\tau)) + \sqrt{b}(g(\tau) - v_n(\tau)) \right) \Big\|.$$

Next collect two facts: Fact 1, uniformly in i

$$\frac{1}{\dot{\lambda}_n^{1/2}} \leq \frac{\Big\| \widehat{V}^{1/2}(\tau) \left(\sqrt{b}(v_{i,b,n}(\tau) - g(\tau)) + \sqrt{b}(g(\tau) - v_n(\tau)) \right) \Big\|}{\Big\| V^{1/2}(\tau) \left(\sqrt{b}(v_{i,b,n}(\tau) - g(\tau)) + \sqrt{b}(g(\tau) - v_n(\tau)) \right) \Big\|} \leq \bar{\lambda}_n^{1/2},$$

where

$$\bar{\lambda}_n = \sup_{\tau} \ \text{maxeig} \ \left(V^{-1/2}(\tau) \widehat{V}(\tau) V^{-1/2}(\tau) \right)$$

and

$$\dot{\lambda}_n = \sup_{\tau} \ \text{maxeig} \ \left(\widehat{V}^{-1/2}(\tau) V(\tau) \widehat{V}^{-1/2}(\tau) \right)$$

by equality 10 on p.460 in Amemiya (1985).[9] Fact 2 follows from Fact 1 and by $|\|A\| - \|w\|| \leq \|A + w\| \leq \|A\| + \|w\|$,

$$1[A_i < (x/u_n - w_n)] \leq 1[\widehat{A}_i < x] \leq 1[A_i < (x/l_n + w_n)]$$

where $l_n = 1/\dot{\lambda}_n^{1/2}$, $u_n = \bar{\lambda}_n^{1/2}$, and w_n is defined below.

By A.2, A.3, and assumptions on $\widehat{V}(\tau)$ and $V(\tau)$

$$w_n \equiv \sup_{\tau} \sqrt{b} \Big\| V^{1/2}(\tau)(v_n(\tau) - g(\tau)) \Big\| = O_p(\sqrt{b}/\sqrt{n}) \xrightarrow{P_n^*} 0,$$
$$q_n \equiv \max[|u_n - 1|, |l_n - 1|] \xrightarrow{P_n^*} 0.$$

[9] $\lambda = \sup_x x'Ax/x'x$ equals the maximum eigenvalue (characteristic root) of the symmetric matrix A.

Thus $1(E_n) = 1$ wp$\to 1$, where $E_n \equiv \{w_n, q_n \leq \delta\}$ for any $\delta > 0$.

II. We have that $E_{P_n}[\dot{G}_{n,b}(x)] = P_n(\mathcal{S}_b \leq x)$ by non-replacement sampling. Conclude that $\dot{G}_{n,b}(x) \xrightarrow{P_n^*} G(x)$ by contiguity and a law of large numbers. In the i.i.d. case, the law of large numbers is that for U-statistics of degree b; and for the time series case, the law of large numbers is the one stated in Theorem 3.2.1 in Politis and Romano and Wolf (1999).

Using I, for small enough $\epsilon > 0$ there is $\delta > 0$, so that by fact 2: $\dot{G}_{n,b}(x-\epsilon)1(E_n) \leq \widehat{G}_{n,b}(x)1(E_n) \leq \dot{G}_{n,b}(x+\epsilon)1(E_n)$, so that with probability tending to one: $\dot{G}_{n,b}(x-\epsilon) \leq \widehat{G}_{n,b}(x) \leq \dot{G}_{n,b}(x+\epsilon)$. Next, note that $x = G^{-1}(1-\alpha)$ is a continuity point by assumption. Pick $\epsilon > 0$ so that $[x-\epsilon, x+\epsilon]$ are continuity points of $G(x)$. For such small enough ϵ, $\dot{G}_{n,b}(x+c) \xrightarrow{P_n^*} G(x+c)$, for $c = \pm\epsilon$, which implies $G(x-\epsilon) - \epsilon \leq \widehat{G}_{n,b}(x) \leq G(x+\epsilon) + \epsilon$ w.p. $\to 1$. Since ϵ can be set arbitrarily small, conclude that $G_{n,b}(x) \xrightarrow{P_n^*} G(x)$.

Convergence of quantiles is implied by the convergence of distribution functions at continuity points.

Coverage results follow from the definition of weak convergence.

Finally, that $\beta > \alpha$ follows by an Anderson's Lemma for Banach spaces, Lemma 3.11.4 in van der Vaart and Wellner (1996).

III. (iv) follows from Theorem 11.1 of Davydov, Lifshits, and Smordina (1998) by A.3. □

Proof of Theorem 2. To show (i): the proof is a direct corollary of Theorem 7.3.1 in Politis, Romano and Wolf (1999).

To show (ii): the convergence of truncated moments follows from the definition of ρ_{BL}.

To show (iii): we use direct arguments. Let

$$\bar{v}_{i,b,n}(\cdot) \equiv v_{i,b,n}(\cdot) - v_n(\cdot), \text{ and } f_K^c(x) \equiv x - f_K(x),$$

and write

$$E_{L_{b,n}}\left[(\sqrt{b}\bar{v}_{i,b,n}(\tau))^m\right] = E_{L_{b,n}}\underbrace{\left[f_K\{(\sqrt{b}\bar{v}_{i,b,n}(\tau))^m\}\right]}_{I} + E_{L_{b,n}}\underbrace{\left[f_K^c\{(\sqrt{b}\bar{v}_{i,b,n}(\tau))^m\}\right]}_{II}.$$

For any given K, the term $E_{Ln,b}I$ converges uniformly in τ to $E_P\left[f_K\{\bar{v}(\tau)^m\}\right]$ by part (ii). Thus, it suffices to show that for any $\epsilon > 0$, there is $K > 0$ such that

$$\text{(a) } \sup_{\tau\in\mathcal{T}} |E_{L_{b,n}}II| \leq \epsilon \text{ wp } \to 1, \text{ and } \qquad \text{(b) } E_P \sup_{\tau\in\mathcal{T}} |f_K^c\{\bar{v}(\tau)^m\}| \leq \epsilon.$$

(b) is immediate by the uniform integrability. To show (a), for any given $\epsilon > 0$ and $\delta > 0$

$$\begin{aligned}
&\overline{\lim}_n P_n^* \left[\sup_{\tau \in \mathcal{T}} \left| E_{L_{b,n}} II \right| > \epsilon \right] \\
&\overset{(1)}{\leq} \overline{\lim}_n P_n^* \left[E_{L_{b,n}} \sup_{\tau \in \mathcal{T}} \left| II \right| > \epsilon \right] \\
&\overset{(2)}{\leq} \overline{\lim}_n \ \left[E_{P_n^*} \sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{b} \bar{v}_{i,b,n}(\tau) \right]^m \right\} \right| \right] \Big/ \epsilon \\
&\leq \overline{\lim}_n \ \left[E_{P_n^*} \sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{b} (v_{i,b,n}(\tau) - g_{P_n}(\tau)) \right]^m \right\} \right| \right. \\
&\qquad\qquad \left. + E_{P_n^*} \sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{b} \left(v_n(\tau) - g_{P_n}(\tau) \right) \right]^m \right\} \right| \right] \Big/ \epsilon \\
&\leq \overline{\lim}_n \ 2 \cdot \max_{J_n \in \{b,n\}} \left[E_{P_n^*} \sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{J}_n \left(v_{J_n}(\tau) - g_{P_n}(\tau) \right) \right]^m \right\} \right| \right] \Big/ \epsilon \\
&\overset{(3)}{\leq} \delta,
\end{aligned}$$

where (1) is by Jensen; (2) is by Markov and

$$E_{P^*} \left[E_{L_{b,n}} \sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{b} \bar{v}_{i,b,n}(\tau) \right]^m \right\} \right| \right] = E_{P^*} \left[\sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{b} \bar{v}_{i,b,n}(\tau) \right]^m \right\} \right| \right],$$

by non-replacement sampling and stationarity; (3) follows by picking K sufficiently large such that

$$\sup_{P \in \{P_n, n \geq n_0\}} \sup_{J_n > n_1} \left[E_{P^*} \sup_{\tau \in \mathcal{T}} \left| f_K^c \left\{ \left[\sqrt{J}_n \left(v_{J_n}(\tau) - g_{P_n}(\tau) \right) \right]^m \right\} \right| \right] \leq \frac{1}{2} \cdot \delta \cdot \epsilon,$$

by the uniform integrability.

To show (iv), define, for $\bar{\nu}(\tau) = \sqrt{b}(v_{i,b,n}(\tau) - v_n(\tau))$,

$$G(x; \tau) = P_L\{\bar{v}(\tau)_j \leq x\} \text{ and } G_n(x; \tau) = P_{L_{b,n}}\{\bar{\nu}(\tau)_j < x\}.$$

Denote by $f_K(y)$ a smooth approximation of the indicator function $1(y < 0)$ that disagrees with $1(y < 0)$ only on the set $K = [0, \delta]$ for $\delta > 0$, such that $y \mapsto f_K(y) \in \mathbf{L}(c(\delta), k(\delta))$. Then by part (i)

$$\sup_{(\tau,x) \in \mathcal{T} \times \mathcal{X}} \left| E_L\{f_K(\bar{v}(\tau)_j - x)\} - E_{L_{b,n}}\{f_K(\bar{\nu}(\tau)_j - x)\} \right| \xrightarrow{P_n^*} 1,$$

where $\mathcal{X}$ is any fixed compact set. Using this fact and that the asymptotic limit $v(\cdot)_j$ is a nondegenerate Gaussian process on the compact set (each coordinate of which, $v(\tau)_j$, has, uniformly in τ, uniformly bounded density, which rules out point masses), a standard smoothing argument delivers: $\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}} |G(x;\tau) - G_n(x;\tau)| \xrightarrow{P_n^*} 0$, which by Lemma 1 implies $\sup_{\tau\in\mathcal{T}} |G_n^{-1}(p;\tau) - G^{-1}(p;\tau)| \xrightarrow{P_n^*} 0$. A similar argument is applied to any sum of two components $v(\tau)_j + v(\tau)_k$. $\square$

This lemma is a simple generalization of a well-known result.

LEMMA 1. *Suppose $\{G_n(x|\tau), \tau \in \mathcal{T}\}$ is a collection of distribution functions indexed by τ, and $G(x|\tau)$ is also a cdf such that $G^{-1}(p|\tau)$ is uniformly continuous in (p,τ), for $(p,\tau) \in \mathcal{P}\times\mathcal{T}$, $\mathcal{P} \subset [0,1]$. Suppose also that*

$$\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}} \left| G_n(x|\tau) - G(x|\tau) \right| \xrightarrow{P_n^*} 0.$$

Then, for any p such that: $\cup_{\tau\in\mathcal{T}}\{G^{-1}(p\pm\delta|\tau)\} \in \mathcal{X}$ for some $\delta > 0$, we have that:

$$\sup_{\tau\in\mathcal{T}} \left| G_n^{-1}(p|\tau) - G^{-1}(p|\tau) \right| \xrightarrow{P_n^*} 0.$$

PROOF. By assumption, wp $\to 1$, uniformly in τ

$$G_n(G^{-1}(p'|\tau)|\tau) - \epsilon < \underbrace{p'}_{G(G^{-1}(p'|\tau)|\tau)} < G_n(G^{-1}(p'|\tau)|\tau) + \epsilon,$$

uniformly in $p' \in \{p, p-\epsilon, p+\epsilon\}$, for some fixed p. Since G_n is cadlag, this inequality implies:

$$G_n^{-1}(p'+\epsilon|\tau) \geq G^{-1}(p'|\tau) \geq G_n^{-1}(p'-\epsilon|\tau).$$

Apply this inequality to $p' \in \{p, p-\epsilon, p+\epsilon\}$, and conclude that wp $\to 1$, uniformly in τ

$$G^{-1}(p+2\epsilon|\tau) \geq G_n^{-1}(p+\epsilon|\tau) \geq G^{-1}(p|\tau) \geq G_n^{-1}(p-\epsilon|\tau) \geq G^{-1}(p-2\epsilon|\tau). \tag{5}$$

Also by monotonicity

$$G_n^{-1}(p+\epsilon|\tau) \geq G_n^{-1}(p|\tau) \geq G_n^{-1}(p-\epsilon|\tau). \tag{6}$$

By assumption we can set ϵ arbitrarily small so that uniformly in τ

$$|G^{-1}(p+2\epsilon|\tau) - G^{-1}(p-2\epsilon|\tau)| < \epsilon. \tag{7}$$

Thus (5)-(7) imply $\sup_{\tau\in\mathcal{T}} \left| G_n^{-1}(p|\tau) - G^{-1}(p|\tau) \right| < \epsilon$ wp $\to 1$. $\square$.

Acknowledgement. We would like to thank Guido Kuersteiner, Geraint Jones, a co-editor, and an anonymous referee for useful comments.

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *J. Amer. Statist. Assoc.*, **97**, 284-292.

Abadie, A., Angrist, J. and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, **70**, 91-117.

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

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2005). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, forthcoming.

Bhattacharya, P.K. (1963). On an analog of regression analysis. *Ann. Math. Statist.*, **34**, 1459-1473.

Buchinsky, M. and Hahn, J. (1998). An alternative estimator for the censored regression model. *Econometrica*, **66**, 653-671.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, **14**, 1171-1179.

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

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73**, 245-261.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, **71**, 1405-1441.

Davydov, Y. A., Lifshits, M.A. and Smorodina, N.V. (1998). *Local Properties of Distributions of Stochastic Functionals.* American Mathematical Society, Providence, RI. Translated from the 1995 Russian original by V. E. Nazălkinskĭ and M.A. Shishkova.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.*, **2**, 267-277.

Firpo, S. (2004). Efficient semiparametric estimation of quantile treatment effects. Working Paper, University of California at Berkeley.

Fukuchi, J.-I. (1999). Subsampling and model selection in time series analysis. *Biometrika*, **86**, 591-604.

Heckman, J., Smith, J. and Clements, N. (1997). Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. *Rev. Econom. Stud.*, **64**, 487-535.

Honoré, B., Khan, S. and Powell, J.L. (2002). Quantile regression under random censoring. *J. Econometrics*, **109**, 67-105.

Khmaladze, É.V. (1988). An innovation approach to goodness-of-fit tests in R^m. *Ann. Statist.*, **16**, 1503-1516.

Koenker, R. and Bassett, G.S. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Koenker, R. and Bilias, Y. (2001). Quantile regression for duration data: a reappraisal of the Pennsylvania reemployment bonus experiments. *Empirical Economics*, to appear, also at `http://www.econ.uiuc.edu/~roger/research/home.html`.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. Preprint, Department of Economics, University of Illinois.

McFadden, D. (1989). Testing for stochastic dominance. In *Studies in Economics of Uncertainty in Honor of Josef Hadar*, T. Fomny, and T. Seo (eds.), Springer, New York.

Meyer, B. (1995). Lesson from the U.S. unemployment insurance experiments. *J. Economic Literature*, **33**, 131-191.

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

Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.

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

Portnoy, S. (1997). Local asymptotics for quantile smoothing splines. *Ann. Statist.*, **25**, 414-434.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.*, **98**, 1001-1012.

Powell, J.L. (1986). Censored regression quantiles. *J. Econometrics*, **32**, 143-155.

Sakov, A. and Bickel, P.J. (2000). An Edgeworth expansion for the m out of n bootstrapped median. *Statist. Probab. Lett.*, **49**, 217-223.

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

Victor Chernozhukov and
Iván Fernández-Val
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive, E52-262 F
Cambridge MA 02142
E-mail: vchern@mit.edu
ifern12@mit.edu

Paper received: September 2004; revised May 2005.