UNIVERSITY OF TORONTO
DEPARTMENT OF ECONOMICS

Working Paper 277

Nonparametric Inferences on Conditional Quantile Processes

Chuan Goh
Department of Economics
University of Toronto
150 Saint George Street
Toronto, Ontario Canada
M5S 3G7

January 2007

Nonparametric Inferences on Conditional Quantile Processes

Chuan Goh*

31 July 2005
Last update: 15 January 2007

Abstract

This paper is concerned with tests of restrictions on the sample path of conditional quantile processes. These tests are tantamount to assessments of lack of fit for models of conditional quantile functions or more generally as tests of how certain covariates affect the distribution of an outcome variable of interest. This paper extends tests of the generalized likelihood ratio (GLR) type as introduced by Fan et al. (2001) to nonparametric inference problems regarding conditional quantile processes. As such, the tests proposed here present viable alternatives to existing methods based on the Khmaladze (1981, 1988) martingale transformation. The range of inference problems that may be addressed by the methods proposed here is wide, and includes tests of nonparametric nulls against nonparametric alternatives as well as tests of parametric specifications against nonparametric alternatives. In particular, it is shown that a class of GLR statistics based on nonparametric additive quantile regressions have pivotal asymptotic distributions given by the suprema of squares of Bessel processes, as in Hawkins (1987) and Andrews (1993). The tests proposed here are also shown to be asymptotically rate-optimal for nonparametric hypothesis testing according to the formulations of Ingster (1993a,b,c) and of Spokoiny (1996).

JEL Classification: C12, C14, C15, C21.

KEYWORDS: Quantile regression, nonparametric inference, minimax rate, additive models, local polynomials, generalized likelihood ratio

*Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario, Canada M5S 3G7. Telephone: +1 (416) 978-4964. Facsimile: +1 (416) 978-6713. goh@economics.utoronto.ca; http://www.chuangoh.org.

1 Introduction

This paper continues a line of work begun in recent years that aims to provide practitioners with tools appropriate for making inferences regarding the sample path of a conditional quantile process. In the first instance, inference problems of this nature are of demonstrated importance in the context of assessments of lack of fit for postulated models of conditional quantile functions. These problems can also be re-expressed as questions relevant to the analysis of treatment effects in biostatistics and in the evaluation of social programmes, where it is often of interest to know how a particular intervention or policy affects the distribution of the outcome variable of interest. In particular, distributional hypotheses associated with these questions can often be reformulated in terms of restrictions on the sample path of the associated conditional quantile process.

Much of the recent econometric work in this area has focused on the development of asymptotic distributional approximations based on empirical quantile regression processes. In the context of linear quantile regressions, Koenker and Machado (1999) introduced goodness-of-fit processes analogous to the least-squares R^2 statistic with the goal of allowing practitioners to test composite hypotheses regarding the effect of covariates over a range of quantiles.

Subsequent research by Koenker and Xiao (2002) investigated the asymptotic distribution theory of "Wald-type" Kolmogorov-Smirnov goodness-of-fit tests appropriate for composite hypotheses regarding the form of linear conditional quantile functions over a range of quantiles.[1] The difficulty considered by Koenker and Xiao (2002) involves the requirement to estimate nuisance parameters in important cases where the researcher is unable to specify completely the form of the conditional quantile process under the null.[2] In such cases the corresponding Kolmogorov-Smirnov test statistic is not asymptotically pivotal, a problem analogous to that considered by Durbin (1973) for the one-sample Kolmogorov-Smirnov test when there exist estimated parameters under the null. Koenker and Xiao (2002) solve the problem posed by the non-pivotal nature of the asymptotic null distribution by adapting an idea of Khmaladze (1981, 1988) in which a trans-

[1]The setting considered in Koenker and Xiao (2002) difers slightly from that considered by Koenker and Machado (1999, §2.3), as the Wald-type inference processes for quantile regression considered in the earlier work did not involve the presence of nuisance parameters under the null.

[2]For example, the researcher may desire a test of the hypothesis that the distribution of the responses under a binary treatment differs from that under the null by a pure location shift at all quantiles in some interval $[\tau, 1-\tau] \subset (0,1)$, but may be unwilling to specify the sign and magnitude of the location shift.

formation of the Kolmogorov-Smirnov statistic results in the elimination of the non-pivotal components of the test statistic, resulting in a goodness-of-fit process equal to a martingale that converges to a standard Brownian motion. The derivation of the Khmaladze transformation is based on the Doob-Meyer decomposition of a parametric empirical quantile process and its practical implementation requires the estimation of nonparametric nuisance functions. [3]

This paper proposes to extend the existing toolkit that has been developed for inference problems regarding conditional quantile processes in two different directions. In particular, an alternative is proposed to existing methods based on Khmaladzation. In addition, this paper also generalizes the inferential domain from families of (approximately) linear conditional quantile functions to certain families of nonparametric conditional quantile functions. It is shown below that tests of the generalized likelihood ratio (GLR) type—first developed in a number of nonparametric mean regression contexts by Fan et al. (2001)—are both feasible and advantageous for nonparametric inference problems involving the sample path of conditional quantile processes.[4] The range of inference problems regarding conditional quantile processes that can be handled in this framework is wide, and includes tests of nonparametric nulls against nonparametric alternatives, as well as tests of parametric specifications of conditional quantile functions against broadly defined nonparametric alternatives.

The GLR test statistics proposed in this paper are shown to have asymptotic null distributions that are free of nuisance parameters. In particular, the limiting

[3]Chernozhukov and Fernández-Val (2005) have recently proposed an alternative to the somewhat complex solution of "Khmaladzation" proposed by Koenker and Xiao (2002) for problems of this type. Noting that the Kolmogorov-Smirnov statistic that is the focus of the paper of Koenker and Xiao (2002) is based on the linear quantile regression process, Chernozhukov and Fernández-Val (2005) proposed to estimate the null distribution of the test statistic by subsampling an appropriately re-centred empirical quantile regression process.

[4]The families of conditional quantile functions considered here are assumed to be members of a smooth class of functions—in particular, function classes associated with the names of Besov, Hölder or Sobolev. As described in Section 2, this paper also explicitly considers families of conditional quantile functions that have the additive structure

$$F^{-1}_{Y_i|\boldsymbol{X}_i}(\alpha) = m_{0,\alpha} + \sum_{j=1}^{d} m_{j,\alpha}(X_{ij}),$$

where X_{ij} is the jth component of $\boldsymbol{X}_i \in \mathbb{R}^d$. As such, the work in this paper can also be viewed as an extension to a quantile regression setting of the GLR inference procedures for nonparametric additive mean regression models proposed by Fan and Jiang (2005).

null distributions are given by the suprema taken over closed subintervals of $(0,1)$ of the squares of scalar-valued Bessel processes, as in De Long (1981); Hawkins (1987) and Andrews (1993). Following Kiefer (1959), the GLR test statistics consequently have a limiting behaviour identical to that of a squared Kolmogorov-Smirnov statistic in a setting without estimated nuisance parameters.

The inference procedures proposed here are also shown to generalize a familiar power-optimality property for parametric likelihood ratio tests in the sense of achieving optimal rates of convergence for nonparametric hypothesis testing according to the formulations of Ingster (1993a,b,c) and of Spokoiny (1996)—this further highlights the usefulness of the methods proposed below.

The remainder of this paper proceeds as follows. Section 2 describes the specific class of nonparametric quantile regression model considered in this paper and provides informal details regarding a natural estimation procedure for members of that class. Section 3 introduces the class of testing problem and test statistic considered here with specific reference to the class of quantile regression model presented in Section 2. Section 4 presents the large-sample properties of the GLR testing procedure proposed here, while Section 5 reports the results of two sets of simulation experiments designed to verify the quality of the asymptotic approximations presented in Section 4. Section 6 concludes. Proofs of theorems and a statement of the regularity conditions presumed to underlie the analysis presented here appear in the appendix.

Notational conventions

All limits are taken as $n \to \infty$, where n denotes the sample size. The symbol $\mathcal{A}$ denotes a set whose closure is contained in $(0,1)$. $D[0,1]$ denotes the space of bounded real-valued right-continuous functions with left-hand limits on [0,1] endowed with the uniform metric. $D\mathcal{A}$ denotes the analogous space defined for cadlag functions on $\mathcal{A}$. The relation $\Rightarrow$ denotes the weak convergence[5] of sequences of measurable random elements in $D[0,1]$ or in $D\mathcal{A}$, while $\xrightarrow{d}$ and $\xrightarrow{p}$ denote convergence in distribution and in probability, respectively. The relation $A_n \overset{p}{=} B_n$ holds if and only if $A_n - B_n \xrightarrow{p} 0$. The assumption of Borel-measurability is maintained throughout for any sequence of random variables or vectors converging in probability to zero. Finally, define for $\alpha \in (0,1)$ the "check function" $\rho_\alpha(u) \equiv u\,\{\alpha - 1(u<0)\}$ and the corresponding score function $\psi_\alpha(u) \equiv \alpha - 1(u<0)$.

[5]Cf. e.g., Billingsley (1999, p. 1).

2 Nonparametric Estimators of Additive Quantile Regressions

We suppose the researcher has a sample of n observations

$$(Y_1, \boldsymbol{X}_1^T)^T, \dots, (Y_n, \boldsymbol{X}_n^T)^T,$$

where the Y_i's are real-valued outcome variables, and the $\boldsymbol{X}_i$'s are d-dimensional covariates for some finite $d \geq 2$. The jth component of $\boldsymbol{X}_i$ is denoted X_{ij}. In this paper, the assumption is maintained that the outcome variables are generated by the quantile regression model

$$Y_i = m_{0,\alpha} + \sum_{j=1}^{d} m_{j,\alpha}(X_{ij}) + \epsilon_{i\alpha}, \tag{1}$$

where for $\alpha \in (0,1)$, $m_{0,\alpha}$ is an unknown constant, $m_{j,\alpha}(\cdot)$ ($j = 1, \dots, d$) are unknown functions, and $\epsilon_{i\alpha}$ is an unobserved random disturbance with conditional α-quantile given $\boldsymbol{X}_i = \boldsymbol{x}$ equal to zero for almost every $\boldsymbol{x}$ in the support of $\boldsymbol{X}_i$.

This paper is concerned with nonparametric inference for the families of conditional quantile functions implied by the model given in (1). The focus here on an additive, rather than a fully nonparametric, quantile regression specification is motivated by a common desire in empirical practice to avoid the small-sample imprecision attendant in fully nonparametric settings brought about by the curse of dimensionality when several covariates are involved. Nonparametric additive modelling involves an effective reduction of the dimension of the statistical problem while naturally retaining a greater degree of flexibility than simple parametric models.[6]

The analysis in this paper proceeds from the assumption that the additive components $m_{1,\alpha}, \dots, m_{d,\alpha}$ in (1) are estimated at the optimal pointwise rate $n^{-\frac{q}{2q+1}}$ when each $m_{j,\alpha}$, for $j = 1, \dots, d$, is q-times continuously differentiable. To the best of this author's knowledge, the only estimator of the additive components in (1) that attains this level of performance is the procedure proposed by Horowitz and Lee (2005).[7] As such, the metaphorical empirical researcher in this paper is assumed to adopt the estimation procedure for (1) set out in Horowitz and Lee

[6]Cf. *inter alia* Hastie and Tibshirani (1990), Fan and Gijbels (1996) and Horowitz and Lee (2002) for examples of how nonparametric additive modelling has been applied in empirical practice.

[7]Other estimators of the nonparametric additive quantile regression model given in (1) include

(2005). The regularity conditions adopted in the analysis presented below are also identical to those adopted by Horowitz and Lee (2005), and for ease of reference are repeated here in Appendix A.

For the sake of completeness, an informal description of the estimation procedure of Horowitz and Lee (2005) is now given.[8] First let x^j denote the jth component of an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^d$, and set $m_\alpha(\boldsymbol{x}) \equiv \sum_{j=1}^d m_{j,\alpha}(x^j)$. Without loss of generality, assume that the support of the covariates in (1) is $S_{\boldsymbol{X}} \equiv [-1,1]^d$, and for each of the additive components in (1), impose the normalization

$$\int_{-1}^{1} m_j(x)dx = 0, \ j = 1, \ldots, d$$

in order to ensure identifiability.

The additive components in (1) are estimated using a two-step procedure. The first step involves the fitting of a series estimator to the conditional α-quantile function implied by (1). In this connection, let $\{p_b : b = 1, 2, \ldots\}$ denote a complete orthonormal basis for smooth functions defined on $[-1,1]$ and satisfying the assumptions detailed below in Appendix A. For a given positive intger B, set

$$P_B(\boldsymbol{x}) \equiv \left[1, p_1(x^1), \ldots, p_B(x^1), p_1(x^2), \ldots, p_B(x^2), \ldots, p_1(x^d), \ldots, p_B(x^d)\right]^T.$$

As such, a series approximation to the conditional α-quantile function $m_{0,\alpha} + m_\alpha(\boldsymbol{x})$ is given by $P_B(\boldsymbol{x})^T \boldsymbol{\theta}_B$, where $\boldsymbol{\theta}_B \in \mathbb{R}^{Bd+1}$. Access to a tractable large-sample theory for the first stage depends on the tuning parameter B satisfying certain conditions given in Appendix A as $n \to \infty$. Assuming these conditions are satisfied, $\boldsymbol{\theta}_B$ is estimated with

$$\hat{\boldsymbol{\theta}}_{n,B} \equiv \arg\min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \rho_\alpha \left(Y_i - P_B(\boldsymbol{X}_i)^T \boldsymbol{\theta}\right).$$

The first-step series estimator of the conditional α-quantile function given $\boldsymbol{x} \in S_{\boldsymbol{X}}$ is accordingly given as

$$\bar{m}_{n,0,\alpha} + \bar{m}_{n,\alpha}(\boldsymbol{x}) \equiv P_B(\boldsymbol{x})^T \hat{\boldsymbol{\theta}}_{n,B},$$

those involving splines (cf. e.g., Doksum and Koo, 2000), backfitting (cf. e.g., Fan and Gijbels, 1996, p. 296–297) and marginal integration (de Gooijer and Zerom, 2003). The large-sample theory of spline and backfitting estimators is complicated to the extent that pointwise rates of convergence and limiting distributions are unknown. On the other hand, the estimator of de Gooijer and Zerom (2003) is asymptotically normal, but takes as a starting point a fully nonparametric d-variate quantile regression model, and consequently suffers from the curse of dimensionality.

[8] Readers already familiar with this estimator might wish to skip ahead to Section 3.

where $\bar{m}_{n,0,\alpha}$ denotes the first component of $\hat{\boldsymbol{\theta}}_{n,B}$ and $\bar{m}_{n,\alpha}(\boldsymbol{x})$ denotes the sum of the individual series estimators $\bar{m}_{n,j,\alpha}(x^j)$ of $m_{j,\alpha}(x^j)$, $j = 1, \ldots, d$. The first-stage series estimator obviously imposes an explicit assumption of additivity, and as such, allows for a faster rate of convergence in bias than would be possible in the case of a fully nonparametric d-variate estimator.

The second stage of the procedure of Horowitz and Lee (2005) involves the sequential fitting of univariate locally polynomial α-quantile regressions for each of the additive components in (1) with the other additive components replaced with the corresponding series estimates obtained in the first stage. In particular, the second-stage estimator of the jth additive component $m_{j,\alpha}(x^j)$ requires an estimator of

$$\begin{aligned} m_{-j,\alpha}(\boldsymbol{X}_{-j,i}) \equiv\ & m_{1,\alpha}(X_{i1}) + \cdots + m_{j-1,\alpha}(X_{i,j-1}) + m_{j+1,\alpha}(X_{i,j+1}) + \cdots \\ & + m_{d,\alpha}(X_{id}), \end{aligned} \tag{2}$$

where $\boldsymbol{X}_{-j,i} \equiv (X_{i1}, \ldots, X_{i,j-1}, X_{i,j+1}, \ldots, X_{id})^T$. Let $\bar{m}_{n,-j,\alpha}(\boldsymbol{X}_{-j,i})$ denote the quantity in (2) with each of the functions on the right-hand side replaced by its corresponding first-stage series estimate. Assuming that $m_{j,\alpha}$ is at least q-times continuously differentiable on $[-1, 1]$, $m_{j,\alpha}(x^j)$ is estimated using $\hat{m}_{n,j,\alpha}(x^j) \equiv \hat{\beta}^j_{n,0,\alpha}$, where

$$\hat{\boldsymbol{\beta}}^j_{n,\alpha} \equiv \left(\hat{\beta}^j_{n,0,\alpha}, \hat{\beta}^j_{n,1,\alpha}, \ldots, \hat{\beta}^j_{n,q-1,\alpha}\right)^T$$

is defined as

$$\begin{aligned} \hat{\boldsymbol{\beta}}^j_{n,\alpha} \equiv\ & \arg \min_{\{\beta^j_0, \beta^j_1, \ldots, \beta^j_{q-1}\}} \frac{1}{nh_n} \sum_{i=1}^{n} \rho_\alpha \left(Y_i - \bar{m}_{n,0,\alpha} - \beta^j_0 - \sum_{k=1}^{q-1} \beta^j_k \left[\frac{X_{ij} - x^j}{h_n}\right]^k \right. \\ & \left. - \bar{m}_{n,-j,\alpha}(\boldsymbol{X}_{-j,i}) \right) \cdot K\left(\frac{x^j - X_{ij}}{h_n}\right), \end{aligned} \tag{3}$$

where $K : [-1, 1] \to \mathbb{R}_+$ and h_n respectively denote a kernel function and bandwidth satisfying conditions detailed in Appendix A. The fitting of the locally polynomial regression in the second stage essentially creates a local averaging effect that reduces the variance of the estimators of the additive components over that attained by the first-stage series estimates—thus it is the second stage that allows each additive component in (1) to be estimated at the optimal rate of convergence. Combining the first-stage estimate of the constant term $m_{0,\alpha}$ and the second-stage estimates of the additive functions yields

$$\hat{m}_{n,\alpha}(\boldsymbol{X}_i) \equiv \bar{m}_{n,0,\alpha} + \sum_{j=1}^{d} \hat{m}_{n,j,\alpha}(x^j)$$

as an estimator of the conditional α-quantile of Y_i given $\boldsymbol{X}_i = \boldsymbol{x}$.

This section concludes by reproducing, for ease of reference, the main result of Horowitz and Lee (2005).

Proposition 1. *Horowitz and Lee (2005, Theorem 3)*

Suppose the conditions set out in Appendix A are valid. Also suppose that the order of continuous differentiability q of each of the additive components in (1) satisfies $q \geq 2$ and is even. Let $j \in \{1, \ldots, d\}$.

Then for any x^j satisfying $|x^j| \leq 1 - h_n$, where h_n is the bandwidth used to implement the second-stage locally polynomial regression,

$$\left|\hat{m}_{n,j,\alpha}(x^j) - m_{j,\alpha}(x^j)\right| = O_p\left(n^{-\frac{q}{2q+1}}\right),$$

and

$$n^{\frac{q}{2q+1}}\left(\hat{m}_{n,j,\alpha}(x^j) - m_{j,\alpha}(x^j)\right) \xrightarrow{d} N\left(B(x^j), V(x^j)\right),$$

where $B(x^j)$ and $V(x^j)$ are as described in the statement of Horowitz and Lee (2005, Theorem 3). Furthermore, for $k \neq j$ and any x^k satisfying $|x^k| \leq 1 - h_n$,

$$n^{\frac{q}{2q+1}}\left(\hat{m}_{n,j,\alpha}(x^j) - m_{j,\alpha}(x^j)\right)$$

and $n^{\frac{q}{2q+1}}\left(\hat{m}_{n,k,\alpha}(x^k) - m_{k,\alpha}(x^k)\right)$ are asymptotically independently normally distributed.

3 The Testing Problem and Test Statistic

The class of inference problem and associated class of test statistic considered in this paper are defined in this section. First note that the class of model given above in (1) implies that the conditional α-quantile function of the outcome variable satisfies

$$F^{-1}_{Y_i|\boldsymbol{X}_i}(\alpha) = m_{0,\alpha} + \sum_{j=1}^{d} m_{j,\alpha}(X_{ij}). \tag{4}$$

As such, the covariates are seen to be capable of potentially affecting the entire shape—and not merely the location—of the conditional distribution of the outcome variable.

The assumption is maintained that the specification in (4) holds for all $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is taken to be a closed subinterval of $(0, 1)$. This paper is concerned with the general hypothesis

$$H_0 : \{m_{j_1,\alpha}, \dots, m_{j_H,\alpha}\} \subset \mathcal{M}_{H_0}(\alpha) \; \forall \alpha \in \mathcal{A}, \tag{5}$$

where for some $1 \leq H \leq d$, $\{j_1, \dots, j_H\} \subset \{1, \dots, d\}$. The set $\mathcal{M}_{H_0}(\alpha)$ in (5) denotes a collection of functions, possibly indexed by a finite-dimensional parameter, that satisfies some restriction of interest for all quantiles α in the index set $\mathcal{A}$. The hypothesis given in (5) is clearly of interest when assessing the fit of a certain parametric specification of the conditional α-quantile function. More generally, (5) nests a broad class of inference problems regarding the effect of the covariate vector on the conditional distribution of the outcome variable. In particular, several hypotheses customarily of interest in the analysis of treatment effects can be seen to fit into the general rubric expressed by (5) when the vector of covariates contains an intervention or policy variable.[9]

It is assumed that estimates of the additive components in (4) satisfying the restrictions of the null hypothesis are available,[10] and that under the null, the restricted estimates also converge at the same pointwise rate as the unrestricted estimates.[11] Denote by $\tilde{m}_{n,1,\alpha}, \dots, \tilde{m}_{n,d,\alpha}$ the restricted estimates of $m_{j,\alpha}$ ($j = 1, \dots, d$) satisfying the constraints of (5). As was the case in Section 2, let $\hat{m}_{n,1,\alpha}, \dots, \hat{m}_{n,d,\alpha}$ denote the unrestricted estimates of the additive components in (4), and let $\bar{m}_{n,0,\alpha}$ denote the first-stage series estimate of the constant term $m_{0,\alpha}$. Let

$$RS_0(\alpha) \equiv \sum_{i=1}^{n} \rho_\alpha \left(Y_i - \bar{m}_{n,0,\alpha} - \sum_{j=1}^{d} \tilde{m}_{n,j,\alpha}(X_{ij}) \right) \tag{6}$$

[9]These hypotheses include tests for the significance of treatment (i.e., non-zero treatment effect) for at least one quantile $\alpha \in \mathcal{A}$; for the homogeneity (i.e., constancy) of treatment effect for all quantiles $\alpha \in \mathcal{A}$ versus the heterogeneity of treatment impact for some $\alpha \in \mathcal{A}$; and the stochastic dominance of treatment for all $\alpha \in \mathcal{A}$. Concrete empirical motivation for the consideration of inferential questions of this nature can be found in e.g., Abadie (2002), Bitler et al. (2006) and Heckman et al. (1997).

[10]For example these restricted estimates might be obtained by solving appropriately constrained variants of the second-stage optimization problem given above in (3).

[11]Cf. the first conclusion in the statement of Proposition 1.

denote the value of the objective function for the nonparametric additive α-quantile regression problem under the restrictions of the hypothesis given in (5). Similarly, let

$$RS_1(\alpha) \equiv \sum_{i=1}^{n} \rho_\alpha \left(Y_i - \bar{m}_{n,0,\alpha} - \sum_{j=1}^{d} \hat{m}_{n,j,\alpha}(X_{ij}) \right) \tag{7}$$

denote the optimized value of the α-quantile regression criterion function. As such, the statistic

$$\lambda_n(H_0, \alpha) \equiv RS_0(\alpha) - RS_1(\alpha) \tag{8}$$

provides a natural assessment of the plausibility of the restrictions of H_0 at a fixed value of $\alpha \in \mathcal{A}$. This paper proposes to test H_0 as given above in (5) by examining values of the statistic

$$\sup_{\alpha \in \mathcal{A}} \left(\frac{\lambda_n(H_0, \alpha) - b_{n,\alpha}(K)}{S_{n,\alpha}} \right)^2, \tag{9}$$

where the normalizing sequences $\{b_{n,\alpha}(K)\}$ and $\{S_{n,\alpha}\}$ are as given below in Section 4.1.

Noting that the statistic $\lambda_n(H_0, \alpha)$ in (8) coincides with the parametric log-likelihood ratio statistic if the disturbance terms in (1) above are homoskedastic and generated by the asymmetric Laplace density $f_\alpha(\epsilon) = \alpha(1-\alpha)\exp\left(-\rho_\alpha(\epsilon)\right)$, this paper follows Fan et al. (2001) by referring to $\lambda_n(H_0, \alpha)$ as a generalized likelihood ratio (GLR) statistic. In what follows, the testing procedure based on the statistic in (9) will be referred to as a GLR test.

4 Main Results

The large-sample theory of the GLR testing procedure proposed above for the general class of hypotheses given in (5) is developed in this section.

4.1 Asymptotic null distribution

First define

$$\mathcal{X}_n \equiv \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}. \tag{10}$$

The assumption is made that the elements of $\mathcal{X}_n$ are supported on the space $S_{\boldsymbol{X}} \equiv [-1, 1]^d$.

Recalling the definitions introduced above of $m_{0,\alpha}$, $\bar{m}_{n,0,\alpha}$, $m_{j,\alpha}(\cdot)$, $\hat{m}_{n,j,\alpha}(\cdot)$ and $\tilde{m}_{n,j,\alpha}(\cdot)$ for $j = 1, \ldots, d$, set $m_\alpha(\boldsymbol{X}_i) \equiv m_{0,\alpha} + \sum_{j=1}^{d} m_{j,\alpha}(X_{ij})$, $\hat{m}_{n,\alpha}(\boldsymbol{X}_i) \equiv \bar{m}_{n,0,\alpha} + \sum_{j=1}^{d} \hat{m}_{n,j,\alpha}(X_{ij})$ and $\tilde{m}_{n,\alpha}(\boldsymbol{X}_i) \equiv \bar{m}_{n,0,\alpha} + \sum_{j=1}^{d} \tilde{m}_{n,j,\alpha}(X_{ij})$. Also define

$$S_{n,\alpha}^2 \equiv \alpha(1-\alpha) \sum_{i=1}^{n} \left(\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{m}_{n,\alpha}(\boldsymbol{X}_i)\right)^2 \tag{11}$$

and

$$\begin{aligned} b_{n,\alpha}(K) \quad \equiv \quad & \frac{1}{2} \left(\alpha(1-\alpha)\right)^{-1} K^d(0) \sum_{i=1}^{n} f_\alpha^3(0|\boldsymbol{X}_i) \left[\left(\tilde{m}_{n,\alpha}(\boldsymbol{X}_i) - m_\alpha(\boldsymbol{X}_i)\right)^2 \right. \\ & \left. - \left(\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - m_\alpha(\boldsymbol{X}_i)\right)^2 \right], \end{aligned} \tag{12}$$

where for any $\boldsymbol{x} \in S_{\boldsymbol{X}}$, $f_\alpha(\epsilon|\boldsymbol{x})$ is the conditional density of the disturbance term in (1) above. For $\alpha \in (0,1)$, let $B(\alpha) \sim N(0, \alpha(1-\alpha))$. Following general usage, we refer to the process $\{Q(\alpha)\}$, where

$$Q(\alpha) \equiv \frac{|B(\alpha)|}{\sqrt{\alpha(1-\alpha)}}, \tag{13}$$

as a Bessel process.

The limiting distribution of the test statistic given above in (9) under H_0 is given as follows.

Theorem 1. *Under the conditions given in Appendix A, and for $S_{n,\alpha}$ and $b_{n,\alpha}(K)$ as defined in (11) and (12) respectively,*

$$\sup_{\alpha \in [\tau, 1-\tau]} \left(\frac{\lambda_n(H_0, \alpha) - b_{n,\alpha}(K)}{S_{n,\alpha}} \right)^2 \Bigg| \mathcal{X}_n \xrightarrow{d} \sup_{\alpha \in [\tau, 1-\tau]} Q^2(\alpha).$$

under H_0 for $\tau \in \left(0, \frac{1}{2}\right]$.

Proof. The proof appears in Appendix B. $\square$

As such, the GLR test statistic behaves asymptotically under the null like the square of a Kolmogorov-Smirnov statistic.[12] Critical values for the limit quantity $\sup_{\alpha \in [\tau, 1-\tau]} Q^2(\alpha)$ can be found in Andrews (1993, Table 1) for a number of different settings of τ.

[12] Cf. Kiefer (1959).

The conclusion of Theorem 1 indicates that the asymptotic null distribution of the test statistic proposed here is free of nuisance parameters. In particular, the unknown quantities in the normalizing constants given above in (11) and (12) can be replaced with reasonable estimates without affecting the asymptotic properties of the test under the null. In addition, any of the additive components appearing in (4) whose values are unaffected by the restrictions of (5) can also be replaced with estimates without affecting the conclusion of Theorem 1. As mentioned above in the Introduction, this property is in contrast to some of the earlier work on this problem, where the asymptotic null distributions of the test statistics are affected by the estimation of nuisance parameters.[13]

4.2 Asymptotic power analysis

This section considers the power in large samples of the GLR testing procedure against a broad class of smooth functional alternatives that shrink to the null as $n \to \infty$. For simplicity, attention is restricted here to the test of the exclusion restriction

$$H_0 : m_{d,\alpha} \equiv 0 \ \forall \alpha \in [\tau, 1-\tau] \tag{14}$$

for some fixed $\tau \in \left(0, \frac{1}{2}\right]$. The assumption is made that for each n, the class of relevant alternatives to the hypothesis in (14) belongs to a Hölder class of differentiable functions on $\mathbb{R}^2$ that vanish as $n \to \infty$. In particular, the sequence of local alternatives $\{H_{1n}\}$ is considered, where

$$H_{1n} : m_{d,\alpha}(x^d) = \mu_n\left(\alpha, x^d\right) \tag{15}$$

for some $\alpha \in [\tau, 1-\tau]$ and all $x^d \in [-1, 1]$, and where $\mu_n\left(\alpha, x^d\right) \to 0$ as $n \to \infty$. Each member of the sequence of functions $\{\mu_n\}$ is assumed to have bounded derivatives of order q on $[\tau, 1-\tau] \times [-1, 1]$.

The local power of the GLR testing procedure for the hypothesis given in (14) against H_{1n} is approximable via the conclusion of the following theorem.

Theorem 2. *Suppose the conditions set out in Appendix A hold, and that H_{1n} in (15) holds for some $\alpha' \in [\tau, 1-\tau]$. Then*

$$\sup_{\alpha \in [\tau, 1-\tau]} \left(\frac{\lambda_n(H_0, \alpha) - b^{(0)}_{n,\alpha}(K)}{S_{n,\alpha}} \right)^2 \Bigg| \mathcal{X}_n \overset{asy}{\sim} \sup_{\alpha \in [\tau, 1-\tau]} Q^2(\alpha) + C^{(1)}_{n,\alpha'}(K),$$

[13] Cf. esp. Koenker and Xiao (2002).

where $S_{n,\alpha}$ is as given above in (11), and where $b^{(0)}_{n,\alpha}(K)$ and $C^{(1)}_{n,\alpha'}(K)$ are as defined below in (20) and (23), respectively.

Proof. The proof appears in Appendix C. □

In a manner analogous to that of Theorem 1, the conclusion of Theorem 2 allows for an assessment of the local power of the GLR testing procedure to be made in simulations with any nuisance parameters replaced with reasonable estimates.

In order to investigate issues of optimality, we define a class of functions $\mathcal{M}_n$ whose members are boundedly differentiable to qth order on $\mathbb{R}^2$ such that for every $\alpha \in [\tau, 1-\tau]$ and every $\mu_n(\alpha, x^d) \in \mathcal{M}_n$,

$$Var\left[f_\alpha(0|\boldsymbol{X}_1)\sigma_\alpha^{-2}(\boldsymbol{X}_1)\mu_n^2(\alpha, X_{1d})\right] < D\left(E\left[f_\alpha(0|\boldsymbol{X}_1)\sigma_\alpha^{-2}(\boldsymbol{X}_1)\mu_n^2(\alpha, X_{1d})\right]\right)^2 \tag{16}$$

for some constant $D > 0$, and where

$$E\left[f_\alpha(0|\boldsymbol{X}_1)\sigma_\alpha^{-2}(\boldsymbol{X}_1)\mu_n^2(\alpha, X_{1d})\right] = O(1).$$

For a given $r > 0$, define

$$\mathcal{M}_n(r) \equiv \left\{\mu_n \in \mathcal{M}_n : \, E\left[f_{\alpha'}(0|\boldsymbol{X}_1)\sigma_{\alpha'}^{-2}(\boldsymbol{X}_1)\mu_n^2(\alpha', X_{1d})\right] \geq r^2\right\}$$

for some $\alpha' \in [\tau, 1-\tau]$, and where α' may in fact depend on n. Denote the critical function of the GLR test of the null given in (14) by ϕ_n. The probability of type-II error of a level-ω test against the alternative

$$H_{1n}(\alpha') : \, m_{d,\alpha'}(x^d) = \mu_n(\alpha', x_d)$$

is denoted by

$$\beta(\omega, \mu_n) \equiv P\left[\phi_n = 0 \,|m_{d,\alpha'} = \mu_n(\alpha', \cdot)\right].$$

Let

$$\beta(\omega, r) \equiv \sup_{\mu_n \in \mathcal{M}_n(r)} \beta(\omega, \mu_n).$$

Following the formulations of Ingster (1993a,b,c),[14] it is desired to find the smallest distance $r_n = r$ such that a test with asymptotically non-trivial power is still possible. Expressed differently, this distance is the distance between the null and the closest class of alternatives that can be detected by the GLR testing procedure proposed here.

In particular, it is desired to find the *minimax rate of testing* r_n.[15] The following theorem derives the minimax rate of testing for the GLR test of (14) against the class of alternatives described by (15).

[14] Also cf. Spokoiny (1996).

[15] In particular, the minimax rate of the level-ω test ϕ_n is the smallest r_n such that

Theorem 3. *Under the conditions of Appendix A, the GLR test of the exclusion restriction given above in (14) can detect alternatives in the class described by (15) at a rate* $r_n = n^{-\frac{q}{4q+2}}$.

Proof. The proof appears in Appendix D. □

We note that the bound $r_n^2 = n^{-\frac{q}{2q+1}}$ coincides with the minimax rate of testing against alternatives lying in Hölder, Sobolev or Besov classes of functions in $\mathbb{R}^2$ with bounded derivatives of order $q \geq 1$.[16] As such, the GLR testing procedure proposed here can be considered to be asymptotically rate-optimal.

5 Numerical Evidence

This section reports the results of two sets of modest Monte Carlo experiments intended to evaluate the small-sample quality of the asymptotic approximations presented in Section 4. In both sets of experiments, 100 replications were used, and the sample size was set to $n = 100$. In addition, both sets of experiments involved estimation of the additive components in the conditional quantile functions according to a reasonable implementation of the two-step method of Horowitz and Lee (2005, §5).[17] The index set of quantiles of interest was taken to be $\mathcal{A} \equiv [.15, .85]$, which for the purpose of computing the test statistic was approximated by a discrete grid $\hat{\mathcal{A}}$ of 100 equally-spaced points covering the interval $[.15, .85]$.

---

1. for every $r > r_n, \omega > 0$ and any $\beta > 0$, there is a constant c such that $\beta(\omega, cr) \leq \beta + o(1)$; and

2. for any sequence $r_n^* = o(r_n)$, there exist $\omega > 0$ and $\beta > 0$ such that for any $c > 0$, $P\left[\phi_n = 1 \,|m_{d,\alpha'} = \mu_n(\alpha', \cdot)\right] = \omega + o(1)$ and $\liminf_n \beta(\omega, cr_n^*) > \beta$.

[16] Cf. Ingster (1982, 1993a,b,c) and Guerre and Lavergne (2002).

[17] In particular, B-splines were used in the first stage, where the number of terms B_n was chosen to minimize the BIC criterion used in Doksum and Koo (2000), namely,

$$n \log \left(\sum_{i=1}^{n} \rho_\alpha \left(Y_i - P_B \left(\boldsymbol{X}_i \right)^T \hat{\boldsymbol{\theta}}_{nB_n} \right) \right) + 2 \left(\log n \right) B_n.$$

A locally linear fit was used for the second stage with a normal density kernel truncated at ± 3.5 and the bandwidth chosen according to the rule of thumb given in Fan and Gijbels (1996, p. 202).

5.1 Size and local power

The first set of experiments involved simulated data generated according to the model

$$Y_i = .75X_{i1} + 1.5\sin\left(.5\pi X_{i2}^2\right) + \left[1 + c\left(X_{i2} + .25X_{i2}^2\right)\right]\epsilon_i, \tag{17}$$

where $\left\{(X_{i1}, X_{i2})^T\right\}$ are generated from a bivariate normal distribution with mean $\mathbf{0}$, unit variance and covariance $.2$, and where ϵ_i is standard normal and independently distributed of the covariates. The constant c in (17) was taken to range over the set $\{0, .2, .4, .6, .8, 1.0\}$. As such, c indexes the distance between the situation where the covariate X_{i2} exerts a "pure location-shift" effect on the conditional distribution of Y_i (i.e., homoskedasticity) and the alternative of X_{i2} having a more complicated location-scale-shift effect. The inference problem considered in this simulation can be accordingly expressed as a test of

$$H_0 : F^{-1}_{Y_i|X_{i1},X_{i2}}(\alpha) = m_{1,\alpha}(X_{i1}) + m_2(X_{i2})\ \forall\alpha \in \mathcal{A} = [.15, .85], \tag{18}$$

where it is emphasized that the function $m_2(\cdot)$ does not vary with α. A quantile-invariant estimator of $m_2(X_{i2})$ was constructed by averaging estimates of the additive component corresponding to X_{i2} over a grid of 100 equally-spaced points $\hat{\mathcal{A}} \equiv \{\alpha_g\}$ covering the interval $[.15, .85]$. This estimate was in turn used along with unrestricted estimates of $m_{1,\alpha_g}(X_{i1})$ in constructing a grid of GLR statistics $\{\lambda_n(H_0, \alpha_g)\}$ for α_g ranging over $\hat{\mathcal{A}}$. Estimates $\hat{S}_{n,\alpha_g}$ and $\hat{b}_{n,\alpha_g}(K)$ of the normalizing constants defined above in (11) and (12), respectively, were constructed for each $\alpha_g \in \hat{\mathcal{A}}$ by substituting reasonable estimates of $m_{\alpha_g}(\boldsymbol{X}_i)$ and $f_{\alpha_g}(0|\boldsymbol{X}_i)$ in the corresponding locations.[18]

Approximating the test statistic

$$\sup_{\alpha\in[.15,.85]} \left(\frac{\lambda_n(H_0,\alpha) - b_{n,\alpha}(K)}{S_{n,\alpha}}\right)^2$$

[18] In particular, an unrestricted estimate $\hat{m}_{n,\alpha_g}(\boldsymbol{X}_i)$ of $m_{\alpha_g}(\boldsymbol{X}_i)$ constructed in the manner described in the introduction to this section was used, while $f_{\alpha_g}(0|\boldsymbol{X}_i)$ was estimated using the difference quotient

$$\hat{f}_{\alpha_g}(0|\boldsymbol{X}_i) \equiv \frac{2\tilde{h}_n}{\hat{m}_{n,\alpha_g+\tilde{h}_n}(\boldsymbol{X}_i) - \hat{m}_{n,\alpha_g-\tilde{h}_n}(\boldsymbol{X}_i)}.$$

Here the smoothing parameter $\tilde{h}_n$ was selected according to the plug-in rule of Hall and Sheather (1988) based on an Edgeworth expansion for studentized quantiles of normally distributed data.

with $\max_{\alpha_g \in \hat{\mathcal{A}}} \left(\frac{\lambda_n(H_0,\alpha_g) - \hat{b}_{n,\alpha_g}(K)}{\hat{S}_{n,\alpha_g}} \right)^2$, rejection frequencies of a 5%-test of H_0 as given in (18) are tabulated in Table 1. The critical value for the test was taken from the appropriate location in Andrews (1993, Table 1) and found to be 8.85. The empirical rejection probabilities across the 100 simulations conducted are suggestive of satisfactory size and power performance in small samples against heteroskedastic local alternatives indexed by the parameter c in (17).

5.2 Sensitivity of the test to estimated nuisance parameters

The goal of the second set of simulations was to assess the sensitivity in small samples of the distribution of the GLR test statistic to variation in nuisance parameters existing under the null. In this connection, the hypothesis of X_{i2} exerting a pure location shift as given above in (18) is maintained. The data-generating process considered here is consistent with the null and is given by the homoskedastic model

$$Y_i = .75X_{i1} + 1.5s \sin\left(.5\pi X_{i2}^2\right) + .25\epsilon_i, \tag{19}$$

where $\{(X_{i1}, X_{i2}, \epsilon_i)\}$ are generated according to the same process used in Section 5.1, and where the constant s is taken to range over the set $\left\{1, \frac{2}{3}, \frac{3}{2}\right\}$. The different settings of s are intended to help evaluate the sensitivity of the behaviour of the test statistic to variation in the (quantile-invariant) nuisance parameter $m_2(X_{i2})$.

Taking $\hat{\mathcal{A}}$ as before to denote a grid of 100 evenly-spaced points covering the interval $[.15, .85]$, a simulated sample of 100 normalized GLR test statistics

$$\max_{\alpha_g \in \hat{\mathcal{A}}} \left(\frac{\lambda_n(H_0,\alpha_g) - b_{n,\alpha_g}(K)}{S_{n,\alpha_g}} \right)^2$$

was obtained for each of the three settings of s considered. In this simulation, the true values of the normalizing constants $b_{n,\alpha_g}(K)$ and S_{n,α_g} for each $\alpha_g \in \hat{\mathcal{A}}$ were used. The densities of the normalized GLR statistics for the location-shift hypothesis were then estimated using a normal density kernel and the rule-of-thumb bandwidth $h_{100} = 1.06\hat{\sigma}_{100}100^{-.2}$, where $\hat{\sigma}_{100}$ is the standard error in simulations of the normalized test statistic. The estimated densities are plotted in Figure 6 for each of the three settings of s considered, and are seen to be very close, as predicted by the conclusion of Theorem 1. As such, Figure 6 suggests that the distribution of the normalized GLR test statistics is indeed pivotal in small samples.

6 Conclusion

This paper has illustrated the utility, feasibility and asymptotic rate-optimality of tests of the generalized likelihood ratio type as developed by Fan et al. (2001) for nonparametric inference regarding conditional quantile processes. As such, the existing set of tools that has been developed to address problems of this type has been extended in a direction permitting an expansion in the range of inferential questions that can be handled by nonparametric quantile-based methods of inference. The GLR tests proposed here also present a viable alternative to existing methods requiring Khmaladzation.

Acknowledgements

This research was supported by the Connaught Fund of the University of Toronto. I am grateful to Christian Gouriéroux for a number of helpful comments. The usual disclaimer applies.

A Assumptions

The analysis in this paper is predicated on the same conditions adopted by Horowitz and Lee (2005). For ease of reference, these conditions are repeated in the following list. Once again, for any vector $\boldsymbol{x} \in \mathbb{R}^d$, x^j denotes the jth component of $\boldsymbol{x}$, while for any matrix $\boldsymbol{M}$, $\|\boldsymbol{M}\| \equiv \sqrt{tr\left[\boldsymbol{M}\boldsymbol{M}^T\right]}$ denotes the Euclidean norm.

1. The researcher is faced with iid data $\left\{(Y_i, \boldsymbol{X}_i^T)^T : i = 1, \ldots, n\right\}$, where the conditional α-quantile of Y_i given $\boldsymbol{X}_i = \boldsymbol{x}$ satisfies

$$F^{-1}_{Y_i|\boldsymbol{X}_i=\boldsymbol{x}}(\alpha) = m_{0,\alpha} + \sum_{j=1}^{d} m_{j,\alpha}(x^j)$$

   for almost every $\boldsymbol{x} \in [-1, 1]^d$.

2. The support of the covariates is given by $S_{\boldsymbol{X}} \equiv [-1, 1]^d$. $\boldsymbol{X}_i$ has an absolutely continuous distribution with respect to Lebesgue measure. The pdf of $\boldsymbol{X}_i$ is bounded, bounded away from zero, is twice continuously-differentiable in the interior of $S_{\boldsymbol{X}}$ and has continuous second-order one-sided derivatives on the boundary of $S_{\boldsymbol{X}}$.

3. Let $F_\alpha(0|\boldsymbol{x})$ denote the conditional distribution function given $\boldsymbol{X}_i = \boldsymbol{x}$ of the error term $\epsilon_{i\alpha}$ in (1). Then $F_\alpha(0|\boldsymbol{x}) = \alpha$ for almost every $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and admits the existence of a (conditional) pdf $f_\alpha(\cdot|\boldsymbol{x})$. There exists a constant $L_\alpha < \infty$ such that $|f_\alpha(e_1|\boldsymbol{x}) - f_\alpha(e_2|\boldsymbol{x})| \leq L_\alpha|e_1 - e_2|$ for all e_1 and e_2 in a neighbourhood of the origin and all $\boldsymbol{x} \in S_{\boldsymbol{X}}$. There also exist constants $c_\alpha > 0$ and $C_\alpha < \infty$ such that $c_\alpha \leq f_\alpha(e|\boldsymbol{x}) \leq C_\alpha$ for all e in a neighbourhood of the origin and all $\boldsymbol{x} \in S_{\boldsymbol{X}}$.

4. For each $j = 1, \ldots, d$, the additive component $m_{j,\alpha}(\cdot)$ is q-times continuously differentiable in the interior of $[-1, 1]$ and has continuous qth-order one-sided derivatives on the boundary of $[-1, 1]$ for some $q \geq 2$.
5. For some positive integer B, set $\mathbf{\Phi}_B \equiv E\left[f_\alpha(0|\boldsymbol{X}_1)P_B(\boldsymbol{X}_1)P_B(\boldsymbol{X}_1)^T\right]$. The smallest eigenvalue of $\mathbf{\Phi}_B$ is bounded away from zero for all B, and the largest eigenvalue of $\mathbf{\Phi}_B$ is bounded for all B.
6. Set $\zeta_B \equiv \sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} \|P_B(\boldsymbol{x})\|$. The basis functions $\{p_b : b = 1, 2, \ldots\}$ satisfy the following:
   (a) Each function p_b is continuous.
   (b) $\int_{-1}^{1} p_b(u)du = 0$.
   (c) $\int_{-1}^{1} p_b(u)p_c(u)du = \begin{cases} 1 & , \quad \text{if } b = c \\ 0 & , \quad \text{otherwise.} \end{cases}$
   (d) As $B \to \infty$, $\zeta_B = O\left(\sqrt{B}\right)$.
   (e) Set $d_B \equiv Bd + 1$. Then there exist vectors $\boldsymbol{\theta}_{B0} \in \mathbb{R}^{d_B}$ such that

   $$\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} \left| m_{0,\alpha} + \sum_{j=1}^{d} m_{j,\alpha}(x^j) - P_B(\boldsymbol{x})^T \boldsymbol{\theta}_{B0} \right| = O\left(B^{-q}\right),$$

   as $B \to \infty$.
7. $\frac{B^4}{n}(\log n)^2 \to 0$ and $\frac{B^{1+2q}}{n}$ is bounded as $B, n \to \infty$.
8. (a) $B = C_B n^\nu$ for some constant $C_B \in (0, \infty)$ and some ν satisfying

   $$\frac{1}{2q+1} < \nu < \frac{2q+3}{12q+6}.$$

   As such, $q \geq 2$.
   (b) $h_n = C_h n^{-\frac{1}{2q+1}}$ for some constant $C_h \in (0, \infty)$.
9. The smoothing kernel $K(\cdot)$ is a bounded continuous and symmetric pdf supported on $[-1, 1]$.
10. Let $\mathbf{0}_B$ denote the zero vector in $\mathbb{R}^B$. For each $j \in \{1, \ldots, d\}$, set

    $$\begin{aligned} \boldsymbol{P}_{B,-j}(\boldsymbol{x}_{-j}) \equiv\ & [1, p_1(x^1), \ldots, p_B(x^1), \ldots, p_1(x^{j-1}), \ldots, p_B(x^{j-1}), \mathbf{0}_B^T, \\ & p_1(x^{j+1}), \ldots, p_B(x^{j+1}), \ldots, p_1(x^d), \ldots, p_B(x^d)]^T, \end{aligned}$$

    where $\boldsymbol{x}_{-j} \equiv (x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^d)^T$. Then the largest eigenvalue of

    $$E\left[\boldsymbol{P}_{B,-j}(\boldsymbol{X}_{1,-j})\boldsymbol{P}_{B,-j}(\boldsymbol{X}_{1,-j})^T \middle| X_{1j} = x^j\right]$$

    is twice continuously differentiable with respect to x^j.

B Proof of Theorem 1

Define the following, for $\delta \in \mathbb{R}$ and $\alpha \in (0,1)$:

$$r_{ni}(\delta,\alpha) \equiv \left[\rho_\alpha\left(\epsilon_{i\alpha} - n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\delta\right) - \rho_\alpha(\epsilon_{i\alpha})\right] + n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\psi_\alpha(\epsilon_{i\alpha})\delta - \frac{1}{2}n^{-\frac{2q}{2q+1}}f_\alpha(0|\boldsymbol{X}_i)K^d(0)\delta^2.$$

and

$$r_n(\delta,\alpha) \equiv \sum_{i=1}^{n} r_{ni}(\delta,\alpha),$$

where, for $\boldsymbol{x} \in \mathcal{S}_{\boldsymbol{X}}$,

$$\sigma_\alpha(\boldsymbol{x}) \equiv \frac{\sqrt{\alpha(1-\alpha)}}{f_\alpha(0|\boldsymbol{x})}.$$

The following preliminary argument is made.

Lemma 1. *For some constant $C > 0$ and $\tau \in \left(0, \frac{1}{2}\right]$, and under the conditions given in Appendix A,*

$$\sup\left\{|r_n(\delta,\alpha)| : |\delta| \le C\sqrt{\log\log n},\ \alpha \in [\tau, 1-\tau]\right\} \xrightarrow{p} 0.$$

Proof. The proof appears in Appendix E.1. □

Now consider $RS_1(\alpha) \equiv \sum_{i=1}^{n} \rho_\alpha(Y_i - \hat{m}_{n,\alpha}(\boldsymbol{X}_i))$. Set

$$\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) \equiv \frac{n^{\frac{q}{2q+1}}\left(\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - m_{n,\alpha}(\boldsymbol{X}_i)\right)}{\sigma_\alpha(\boldsymbol{X}_i)}.$$

Exploiting the conclusion of Lemma 1, we have that

$$\begin{aligned} RS_1(\alpha) &= \sum_{i=1}^{n} \rho_\alpha\left(\epsilon_{i\alpha} - (\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - m_{n,\alpha}(\boldsymbol{X}_i))\right) \\ &= \sum_{i=1}^{n} \rho_\alpha\left(\epsilon_{i\alpha} - n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i)\right) \\ &= \sum_{i=1}^{n}\left[\rho_\alpha(\epsilon_{i\alpha}) - n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\psi_\alpha(\epsilon_{i\alpha})\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) + \frac{1}{2}n^{-\frac{2q}{2q+1}}f_\alpha(0|\boldsymbol{X}_i)K^d(0)\hat{\delta}^2_{n,\alpha}(\boldsymbol{X}_i)\right] \\ &\quad + o_p(1) \end{aligned}$$

uniformly for $\alpha \in [\tau, 1-\tau]$.

Similarly, for

$$\tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) \equiv \frac{n^{\frac{q}{2q+1}}\left(\tilde{m}_{n,\alpha}(\boldsymbol{X}_i) - m_{n,\alpha}(\boldsymbol{X}_i)\right)}{\sigma_\alpha(\boldsymbol{X}_i)},$$

$$RS_0(\alpha) = \sum_{i=1}^{n}\left[\rho_\alpha(\epsilon_{i\alpha}) - n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\psi_\alpha(\epsilon_{i\alpha})\tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) + \frac{1}{2}n^{-\frac{2q}{2q+1}}f_\alpha(0|\boldsymbol{X}_i)K^d(0)\tilde{\delta}^2_{n,\alpha}(\boldsymbol{X}_i)\right] + o_p(1),$$

where the representation holds uniformly for $\alpha \in [\tau, 1-\tau]$.

Then

$$\begin{aligned}
\lambda_n(H_0,\alpha) &\equiv RS_0(\alpha) - RS_1(\alpha) \\
&= \sum_{i=1}^{n} \Big[n^{-\frac{q}{2q+1}} \sigma_\alpha(\boldsymbol{X}_i)\psi_\alpha(\epsilon_{i\alpha}) \left(\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) \right) \\
&\quad + \frac{1}{2} n^{-\frac{2q}{2q+1}} f_\alpha(0|\boldsymbol{X}_i) K^d(0) \left(\tilde{\delta}^2_{n,\alpha}(\boldsymbol{X}_i) - \hat{\delta}^2_{n,\alpha}(\boldsymbol{X}_i) \right) \Big] + o_p(1) \\
&= n^{-\frac{q}{2q+1}} \sum_{i=1}^{n} \left[\sigma_\alpha(\boldsymbol{X}_i) \left(\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) \right) \psi_\alpha(\epsilon_{i\alpha}) \right] \\
&\quad + \frac{1}{2} n^{-\frac{2q}{2q+1}} K^d(0) \sum_{i=1}^{n} f_\alpha(0|\boldsymbol{X}_i) \left(\tilde{\delta}^2_{n,\alpha}(\boldsymbol{X}_i) - \hat{\delta}^2_{n,\alpha}(\boldsymbol{X}_i) \right) + o_p(1).
\end{aligned}$$

The argument is made that the GLR statistic $\lambda_n(H_0, \alpha)$ is conditionally asymptotically normal given $\mathcal{X}_n$.

Lemma 2. *Under the conditions given in Appendix A, and where $S_{n,\alpha}$ and $b_{n,\alpha}(K)$ are as defined above in (11) and (12) respectively,*

$$\left.\frac{\lambda_n(H_0,\alpha) - b_{n,\alpha}(K)}{S_{n,\alpha}}\right| \mathcal{X}_n \xrightarrow{d} N(0,1)$$

under H_0 for $\alpha \in [\tau, 1-\tau]$.

Proof. The proof appears in Appendix E.2. □

Note that the uniform tightness of the sequence

$$\left\{ n^{-\frac{q}{2q+1}} \sum_{i=1}^{n} \sigma_\alpha(\boldsymbol{X}_i) \left(\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) \right) \psi_\alpha(\epsilon_{i\alpha}) \right\}$$

in $D[0,1]$ can be deduced via an appropriate modification of Shorack (1979, Theorem 2.2).

As such, the sequence $\left\{ \frac{\lambda_n(H_0,\alpha) - b_{n,\alpha}(K)}{S_{n,\alpha}} \right\}$ is also uniformly tight in $D[0,1]$.

Uniform tightness and the conclusion of Lemma 2 imply the desired conclusion.

C Proof of Theorem 2

As was the case above, define for $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and $\alpha \in [\tau, 1-\tau]$ the quantity

$$\sigma_\alpha(\boldsymbol{x}) \equiv \frac{\sqrt{\alpha(1-\alpha)}}{f_\alpha(0|\boldsymbol{x})}.$$

Let

$$b_{n,\alpha}^{(0)}(K) \equiv \frac{1}{2}K^d(0)\sum_{i=1}^{n} f_\alpha(0|\boldsymbol{X}_i)\sigma_\alpha^{-2}(\boldsymbol{X}_i)\left\{\left[\sum_{j=1}^{d-1}(\tilde{m}_{n,,j,\alpha}(X_{ij}) - m_{j,\alpha}(X_{ij}))\right]^2 - [\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - m_\alpha(\boldsymbol{X}_i)]^2\right\}. \tag{20}$$

Clearly, under H_0, $b_{n,\alpha}^{(0)}(K) = b_{n,\alpha}(K)$, where $b_{n,\alpha}(K)$ is as given above in (12). In particular, we have that under H_0,

$$\lambda_n(H_0,\alpha) - b_{n,\alpha}^{(0)}(K) = A_{n,\alpha}\ \forall\alpha \in [\tau, 1-\tau],$$

where

$$A_{n,\alpha} \equiv \sum_{i=1}^{n}\psi_\alpha(\epsilon_{i\alpha})\sum_{j=1}^{d-1}(\hat{m}_{n,j,\alpha}(X_{ij}) - \tilde{m}_{n,j,\alpha}(X_{ij})) + \sum_{i=1}^{n}\hat{m}_{n,d,\alpha}(X_{id})\psi_\alpha(\epsilon_{i\alpha}). \tag{21}$$

On the other hand, under H_{1n}, there exists $\alpha' \in [\tau, 1-\tau]$ such that

$$\lambda_n(H_0,\alpha') - b_{n,\alpha'}^{(0)}(K) = A_{n,\alpha'} + B_{n,\alpha'}^{(1)}(K),$$

where

$$B_{n,\alpha'}^{(1)}(K) \equiv \frac{1}{2}K^d(0)\sum_{i=1}^{n} f_{\alpha'}(0|\boldsymbol{X}_i)\sigma_{\alpha'}^{-2}(\boldsymbol{X}_i)\left[\mu_n^2(\alpha', X_{id}) - 2\mu_n(\alpha', X_{id})\sum_{j=1}^{d-1}(\tilde{m}_{n,j,\alpha'}(X_{ij}) - m_{j,\alpha'}(X_{ij}))\right]. \tag{22}$$

It follows that under H_{1n} there exists $\alpha' \in [\tau, 1-\tau]$ such that

$$\left(\frac{\lambda_n(H_0,\alpha') - b_{n,\alpha'}^{(0)}(K)}{S_{n,\alpha'}}\right)^2 = \left(\frac{A_{n,\alpha'}}{S_{n,\alpha'}}\right)^2 + C_{n,\alpha'}^{(1)}(K),$$

where

$$C_{n,\alpha'}^{(1)}(K) \equiv \frac{2A_{n,\alpha'}B_{n,\alpha'}^{(1)}(K) + B_{n,\alpha'}^{(1)^2}(K)}{S_{n,\alpha'}^2}. \tag{23}$$

Therefore under H_{1n}, there exists an $\alpha' \in [\tau, 1-\tau]$ such that

$$\sup_{\alpha\in[\tau,1-\tau]}\left(\frac{\lambda_n(H_0,\alpha) - b_{n,\alpha}^{(0)}(K)}{S_{n,\alpha}}\right)^2 - C_{n,\alpha'}^{(1)}(K) = \sup_{\alpha\in[\tau,1-\tau]}\left(\frac{A_{n,\alpha}}{S_{n,\alpha}}\right)^2.$$

The desired conclusion follows from arguments made above in the proof of Theorem 1.

D Proof of Theorem 3

Let $\bar{q}_{1-\omega}$ denote the $(1-\omega)$-quantile of the distribution of $\sup_{\alpha\in[\tau,1-\tau]} Q^2(\alpha)$. Under H_{1n} as described above in (15), there exists an $\alpha' \in [\tau, 1-\tau]$ such that the probability of type-II error of a level-ω test satisfies

$$\begin{aligned}
&\beta(\omega,\mu_n) \\
= \;& P\left[\sup_{\alpha\in[\tau,1-\tau]}\left(\frac{\lambda_n(H_0,\alpha)-b^{(0)}_{n,\alpha}(K)}{S_{n,\alpha}}\right)^2 < \bar{q}_{1-\omega}\middle| m_{d,\alpha'}=\mu_n(\alpha',\cdot)\right] \\
\leq \;& P\left[\left(\frac{\lambda_n(H_0,\alpha')-b^{(0)}_{n,\alpha'}(K)}{S_{n,\alpha'}}\right)^2 < \bar{q}_{1-\omega}\middle| m_{d,\alpha'}=\mu_n(\alpha',\cdot)\right] \\
= \;& P\left[\left(\frac{A_{n,\alpha'}}{S_{n,\alpha'}}\right)^2 + C^{(1)}_{n,\alpha'}(K) < \bar{q}_{1-\omega}\middle| m_{d,\alpha'}=\mu_n(\alpha',\cdot)\right], \qquad (24)
\end{aligned}$$

where $S_{n,\alpha}$, $b^{(0)}_{n,\alpha}(K)$, $A_{n,\alpha'}$ and $C^{(1)}_{n,\alpha'}(K)$ are as defined in (11), (20), (21) and (23), respectively.

It is clear from (24) that $\beta(\omega,\mu_n)\to 0$ when $C^{(1)}_{n,\alpha'}(K)\to+\infty$, which from (23) occurs when $B^{(1)}_{n,\alpha'}(K)\to+\infty$, where $B^{(1)}_{n,\alpha'}(K)$ is as given above in (22).

Note that

$$\begin{aligned}
\left|B^{(1)}_{n,\alpha'}(K)\right| \;\geq\;& \frac{1}{2}K^d(0)\sum_{i=1}^n f_{\alpha'}(0|\boldsymbol{X}_i)\sigma^{-2}_{\alpha'}(\boldsymbol{X}_i)\mu_n^2(\alpha',X_{id}) \\
&-K^d(0)\sum_{i=1}^n f_{\alpha'}(0|\boldsymbol{X}_i)\sigma^{-2}_{\alpha'}(\boldsymbol{X}_i)\left|\mu_n(\alpha',X_{id})\right| \\
&\cdot\sum_{j=1}^{d-1}\left|\tilde{m}_{n,j,\alpha'}(X_{ij})-m_{j,\alpha'}(X_{ij})\right|. \qquad (25)
\end{aligned}$$

Since $\sum_{j=1}^{d-1}\left|\tilde{m}_{n,j,\alpha'}(X_{ij})-m_{j,\alpha'}(X_{ij})\right| = O_p\left(n^{-\frac{q}{2q+1}}\right)$ from the first conclusion of Proposition 1, and $|\mu_n(\alpha',X_{id})| = o_p(1)$, the second term in (25) satisfies

$$K^d(0)\sum_{i=1}^n f_{\alpha'}(0|\boldsymbol{X}_i)\sigma^{-2}_{\alpha'}(\boldsymbol{X}_i)\left|\mu_n(\alpha',X_{id})\right|\sum_{j=1}^{d-1}\left|\tilde{m}_{n,j,\alpha'}(X_{ij})-m_{j,\alpha'}(X_{ij})\right| = O_p\left(n^{\frac{q+1}{2q+1}}\right).$$

As such, the condition

$$n^{\frac{-q-1}{2q+1}}\sum_{i=1}^n f_{\alpha'}(0|\boldsymbol{X}_i)\sigma^{-2}_{\alpha'}(\boldsymbol{X}_i)\mu_n^2(\alpha',X_{id})\to+\infty$$

with probability approaching one is sufficient for $B^{(1)}_{n,\alpha'}(K)$ to be unbounded in probability. In this connection, consider that

$$\frac{1}{n}\sum_{i=1}^n f_{\alpha'}(0|\boldsymbol{X}_i)\sigma^{-2}_{\alpha'}(\boldsymbol{X}_i)\mu_n^2(\alpha',X_{id}) \overset{p}{=} E\left[f_{\alpha'}(0|\boldsymbol{X}_i)\sigma^{-2}_{\alpha'}(\boldsymbol{X}_i)\mu_n^2(\alpha',X_{id})\right]$$

by the assumption summarized above in (16) that $Var\left[f_{\alpha'}(0|\boldsymbol{X}_1)\sigma_{\alpha'}^{-2}(\boldsymbol{X}_1)\mu_n^2(\alpha', X_{1d})\right]$ is bounded and by Kolmogorov's "first inequality".[19] Therefore

$$\begin{aligned} n^{\frac{-q-1}{2q+1}}\sum_{i=1}^{n} f_{\alpha'}(0|\boldsymbol{X}_i)\sigma_{\alpha'}^{-2}(\boldsymbol{X}_i)\mu_n^2(\alpha', X_{id}) &\overset{p}{=} n^{\frac{q}{2q+1}} E\left[f_{\alpha'}(0|\boldsymbol{X}_1)\sigma_{\alpha'}^{-2}(\boldsymbol{X}_1)\mu_n^2(\alpha', X_{1d})\right] \\ &\geq n^{\frac{q}{2q+1}} r_n^2. \end{aligned}$$

If $n^{\frac{q}{2q+1}} r_n^2 \to \infty$, the smallest value of r_n possible in this setting is $r_n = n^{-\frac{q}{4q+2}}$.

E Proofs of Lemmas

E.1 Proof of Lemma 1

Let $d_{ni\alpha} \equiv n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\delta = n^{-\frac{q}{2q+1}}\frac{\sqrt{\alpha(1-\alpha)}}{f_\alpha(0|\boldsymbol{x})}\delta$. We have

$$\begin{aligned} & r_{ni}(\delta,\alpha) + \frac{1}{2}n^{-\frac{2q}{2q+1}} f_\alpha(0|\boldsymbol{X}_i)K^d(0)\delta^2 \\ = & \left[\rho_\alpha\left(\epsilon_{i\alpha} - d_{ni\alpha}\right) - \rho_\alpha(\epsilon_{i\alpha})\right] + n^{-\frac{q}{2q+1}}\sigma_\alpha(\boldsymbol{X}_i)\psi_\alpha(\epsilon_{i\alpha})\delta \\ = & \left[\left(\epsilon_{i\alpha} - d_{ni\alpha}\right)\left(\alpha - 1\left(\epsilon_{i\alpha} < d_{ni\alpha}\right)\right) - \epsilon_{i\alpha}\left(\alpha - 1\left(\epsilon_{i\alpha} < 0\right)\right)\right] + d_{ni\alpha}\left(\alpha - 1\left(\epsilon_{i\alpha} < 0\right)\right) \\ = & \left[\left(\epsilon_{i\alpha} - d_{ni\alpha}\right)\alpha - \left(\epsilon_{i\alpha} - d_{ni\alpha}\right)1\left(\epsilon_{i\alpha} < d_{ni\alpha}\right) - \epsilon_{i\alpha}\alpha + \epsilon_{i\alpha}1\left(\epsilon_{i\alpha} < 0\right) + d_{ni\alpha}\alpha - d_{ni\alpha}1\left(\epsilon_{i\alpha} < 0\right)\right] \\ = & \left[\left(d_{ni\alpha} - \epsilon_{i\alpha}\right)1\left(0 \leq \epsilon_{i\alpha} < d_{ni\alpha}\right) + \left(\epsilon_{i\alpha} - d_{ni\alpha}\right)1\left(d_{ni\alpha} \leq \epsilon_{i\alpha} < 0\right)\right] \\ \leq & \; 2\left|d_{ni\alpha}\right| \\ = & \; O\left(n^{-\frac{q}{2q+1}}\sqrt{\log\log n}\right). \end{aligned}$$

If $d_{ni\alpha} > 0$, then

$$E\left[r_{ni}(\delta,\alpha)|\,\boldsymbol{X}_i\right] = -\frac{1}{2}n^{-\frac{2q}{2q+1}} f_\alpha(0|\boldsymbol{X}_i)K^d(0)\delta^2 + \int_0^{d_{ni\alpha}}\left(d_{ni\alpha} - u\right)f_\alpha(u|\boldsymbol{X}_i)du. \tag{26}$$

If $d_{ni\alpha} \leq 0$, then

$$E\left[r_{ni}(\delta,\alpha)|\,\boldsymbol{X}_i\right] = -\frac{1}{2}n^{-\frac{2q}{2q+1}} f_\alpha(0|\boldsymbol{X}_i)K^d(0)\delta^2 + \int_{d_{ni\alpha}}^{0}\left(u - d_{ni\alpha}\right)f_\alpha(u|\boldsymbol{X}_i)du. \tag{27}$$

Consider the situation in (26):

$$\begin{aligned} E\left[r_{ni}(\delta,\alpha)|\boldsymbol{X}_i\right] &= -\frac{1}{2}n^{-\frac{2q}{2q+1}} f_\alpha(0|\boldsymbol{X}_i)K^d(0)\delta^2 + O\left(d_{ni\alpha}^2\right) \\ &= -\frac{1}{2}n^{-\frac{2q}{2q+1}} f_\alpha(0|\boldsymbol{X}_i)K^d(0)\delta^2 + O\left(n^{-\frac{2q}{2q+1}}\delta^2\right) \\ &= O\left(n^{-\frac{2q}{2q+1}}\log\log n\right). \end{aligned}$$

[19] Cf. Chung (2001, Theorem 5.3.1).

Similarly, proceeding from (27) also leads to

$$E\left[r_{ni}(\delta,\alpha)|\boldsymbol{X}_i\right] = O\left(n^{-\frac{2q}{2q+1}}\log\log n\right).$$

Now if $d_{ni\alpha} > 0$,

$$\begin{aligned} E\left[r_{ni\alpha}^2(\delta,\alpha)|\boldsymbol{X}_i\right] &\leq \int_0^{d_{ni\alpha}} \left(d_{ni\alpha} - u\right)^2 f_\alpha(u|\boldsymbol{X}_i)du \\ &= O\left(d_{ni\alpha}^3\right) \\ &= O\left(n^{-\frac{3q}{2q+1}}\left(\log\log n\right)^{\frac{3}{2}}\right). \end{aligned}$$

If $d_{ni\alpha} \leq 0$, a similar bound arises:

$$\begin{aligned} E\left[\left. r_{ni}^2(\delta,\alpha)\right| \boldsymbol{X}_i\right] &\leq \int_{d_{ni\alpha}}^0 \left(d_{ni\alpha} - u\right)^2 f_\alpha(u|\boldsymbol{X}_i)du \\ &= O\left(n^{-\frac{3q}{2q+1}}\left(\log\log n\right)^{\frac{3}{2}}\right). \end{aligned}$$

Therefore $Var\left[\left. r_{ni}(\delta,\alpha)\right| \boldsymbol{X}_i\right] = O\left(n^{-\frac{3q}{2q+1}}\left(\log\log n\right)^{\frac{3}{2}}\right)$. It follows that

$$E\left[\left.\sum_{i=1}^n r_{ni}(\delta,\alpha)\right| \mathcal{X}_n\right] = O\left(n^{\frac{1}{2q+1}}\log\log n\right)$$

and

$$Var\left[\left.\sum_{i=1}^n r_{ni}(\delta,\alpha)\right| \mathcal{X}_n\right] = O\left(n^{\frac{-q+1}{2q+1}}\left(\log\log n\right)^{\frac{3}{2}}\right) = O\left(n^{-\frac{1}{2q+1}}\log\log n\right).$$

Recall the definition $r_n(\delta,\alpha) \equiv \sum_{i=1}^n r_{ni}(\delta,\alpha)$. Consider Markov's inequality applied in this context as

$$P\left[\left.|r_n(\delta,\alpha)| \geq s_n\right| \mathcal{X}_n\right] \leq \exp(-s_n t)E\left[\left.\exp\left(r_n(\delta,\alpha)t\right)\right| \mathcal{X}_n\right], \qquad (28)$$

where the assumption is made that $t \in \left(0, n^{\frac{q}{2q+1}}\left(\log\log n\right)^{-\frac{1}{2}}\right)$, thus ensuring that $r_n(\delta,\alpha)t$ is bounded. Set

$$\begin{aligned} M_{r_n}(t) &\equiv E\left[\left.\exp\left(r_n(\delta,\alpha)t\right)\right| \mathcal{X}_n\right] \\ &= \prod_{i=1}^n E\left[\left.\exp\left(r_{ni}(\delta,\alpha)t\right)\right| \boldsymbol{X}_i\right]. \end{aligned}$$

Arguing via Taylor's theorem, there exists a constant $c > 0$ such that

$$\log M_{r_n}(t) \leq tE\left[\left. r_n(\delta,\alpha)\right| \mathcal{X}_n\right] + ct^2 Var\left[\left. r_n(\delta,\alpha)\right| \mathcal{X}_n\right]. \qquad (29)$$

Substituting (29) into (28) produces the bound

$$P\left[\left.|r_n(\delta,\alpha)| \geq s_n\right| \mathcal{X}_n\right] \leq \exp\left(-s_n t + tE\left[\left. r_n(\delta,\alpha)\right| \mathcal{X}_n\right] + ct^2 Var\left[\left. r_n(\delta,\alpha)\right| \mathcal{X}_n\right]\right). \qquad (30)$$

For any $\eta > 0$, set

$$s_n \equiv \eta, \tag{31}$$

and

$$t \equiv n^{-\frac{2}{2q+1}} \log n. \tag{32}$$

Substituting (31) and (32) into (30), we find that there exists a constant n_0 not depending on δ or on α such that

$$P\left[|r_n(\delta, \alpha)| \geq \eta | \mathcal{X}_n\right] \leq n^{-\eta} \tag{33}$$

for all $n \geq n_0$.

A chaining argument is now made in order to complete the proof. For $\zeta, \theta > 0$, let $\alpha_i \equiv \tau + in^{-\frac{\zeta}{2q+1}}$ for $i = 0, 1, 2, \ldots, (1-2\tau)\, n^{\frac{\zeta}{2q+1}}$, and let $\delta_i \equiv -C\sqrt{\log\log n} + in^{-\frac{\theta}{2q+1}}$ for $i = 0, 1, 2, \ldots, 2Cn^{\frac{\theta}{2q+1}}\sqrt{\log\log n}$. Set

$$S \equiv \left\{(\alpha, \delta) : \alpha = \alpha_i,\, \delta = \delta_j \text{ for some } i \text{ and } j\right\}.$$

The cardinality of S is accordingly $O\left(n^{\frac{\zeta+\theta}{2q+1}}\sqrt{\log\log n}\right)$ and we can exploit (33) to get

$$P\left[\sup_S |r_n(\delta, \alpha)| \geq \eta \middle| \mathcal{X}_n\right] \leq n^{\frac{\zeta+\theta}{2q+1}}\sqrt{\log\log n} \cdot n^{-\eta} \tag{34}$$

for n sufficiently large. Now pick $\{\delta_1, \delta_2\} \subset \left\{\delta : |\delta| \leq C\sqrt{\log\log n}\right\}$ with $|\delta_1 - \delta_2| \leq n^{-\frac{\theta}{2q+1}}$ and $\{\alpha_1, \alpha_2\} \subset [\tau, 1-\tau]$ with $|\alpha_1 - \alpha_2| \leq n^{-\frac{\zeta}{2q+1}}$. Set

$$\Delta_i \equiv |r_{ni}(\delta_1, \alpha_1)|,$$

and

$$\Delta \equiv \left|\sum_{i=1}^{n} (r_{ni}(\delta_1, \alpha_1) - r_{ni}(\delta_2, \alpha_2))\right|.$$

Also set

$$d_{ni\alpha_1} \equiv n^{-\frac{q}{2q+1}} \sigma_{\alpha_1}(\boldsymbol{X}_i)\delta_1,$$

and

$$d_{ni\alpha_2} \equiv n^{-\frac{q}{2q+1}} \sigma_{\alpha_2}(\boldsymbol{X}_i)\delta_2.$$

Then

$$\begin{aligned}
\Delta_i = \; & \big|\{(d_{ni\alpha_1} - \epsilon_{i\alpha_1})1\,(0 \leq \epsilon_{i\alpha_1} < d_{ni\alpha_1}) + (\epsilon_{i\alpha_1} - d_{ni\alpha_1})1\,(d_{ni\alpha_1} \leq \epsilon_{i\alpha_1} < 0)\} \\
& - \{(d_{ni\alpha_2} - \epsilon_{i\alpha_2})1\,(0 \leq \epsilon_{i\alpha_2} < d_{ni\alpha_2}) + (\epsilon_{i\alpha_2} - d_{ni\alpha_2})1\,(d_{ni\alpha_2} \leq \epsilon_{i\alpha_2} < 0)\} \\
& - \frac{1}{2} n^{-\frac{2q}{2q+1}} f_{\alpha_1}(0|\boldsymbol{X}_i) K^d(0) \delta_1^2 + \frac{1}{2} n^{-\frac{2q}{2q+1}} f_{\alpha_2}(0|\boldsymbol{X}_i) K^d(0) \delta_2^2 \Big|.
\end{aligned}$$

First consider

$$\begin{aligned}
\Delta_{i1} \equiv \; & \{(d_{ni\alpha_1} - \epsilon_{i\alpha_1})1\,(0 \leq \epsilon_{i\alpha_1} < d_{ni\alpha_1}) + (\epsilon_{i\alpha_1} - d_{ni\alpha_1})1\,(d_{ni\alpha_1} \leq \epsilon_{i\alpha_1} < 0)\} \\
& - \{(d_{ni\alpha_2} - \epsilon_{i\alpha_2})1\,(0 \leq \epsilon_{i\alpha_2} < d_{ni\alpha_2}) + (\epsilon_{i\alpha_2} - d_{ni\alpha_2})1\,(d_{ni\alpha_2} \leq \epsilon_{i\alpha_2} < 0)\}.
\end{aligned}$$

We say that $\Delta_{i1} \in \left\{\Delta_{i1}^{11}, \Delta_{i1}^{01}, \Delta_{i1}^{10}, \Delta_{i1}^{00}\right\}$, where

$$\Delta_{i1} = \begin{cases} \Delta_{i1}^{11} & \text{iff} \quad d_{ni\alpha_1} > 0 \,\&\, d_{ni\alpha_2} > 0; \\ \Delta_{i1}^{01} & \text{iff} \quad d_{ni\alpha_1} \leq 0 \,\&\, d_{ni\alpha_2} > 0; \\ \Delta_{i1}^{10} & \text{iff} \quad d_{ni\alpha_1} > 0 \,\&\, d_{ni\alpha_2} \leq 0; \\ \Delta_{i1}^{00} & \text{iff} \quad d_{ni\alpha_1} \leq 0 \,\&\, d_{ni\alpha_2} \leq 0. \end{cases}$$

We have

$$\begin{aligned}
\Delta_{i1}^{11} &\leq |d_{ni\alpha_1} - d_{ni\alpha_2}| + |\epsilon_{i\alpha_2} - \epsilon_{i\alpha_1}| \\
&= n^{-\frac{q}{2q+1}} \left| \frac{\sqrt{\alpha_1(1-\alpha_1)}}{f_{\alpha_1}(0|\boldsymbol{X}_i)}(\delta_1 - \delta_2) + \delta_2 \left(\frac{\sqrt{\alpha_1(1-\alpha_1)}}{f_{\alpha_1}(0|\boldsymbol{X}_i)} - \frac{\sqrt{\alpha_2(1-\alpha_2)}}{f_{\alpha_2}(0|\boldsymbol{X}_i)} \right) \right| + O\left(n^{-\frac{\zeta}{2q+1}}\right) \\
&\leq O\left(n^{\frac{-q-\theta}{2q+1}}\right) + O\left(n^{\frac{-q-\zeta}{2q+1}} \sqrt{\log\log n}\right) + O\left(n^{-\frac{\zeta}{2q+1}}\right) \\
&= O\left(n^{\frac{-q-\theta}{2q+1}}\right) + O\left(n^{-\frac{\zeta}{2q+1}}\right).
\end{aligned}$$

Similarly,

$$\Delta_{i1}^{00} \leq O\left(n^{\frac{-q-\theta}{2q+1}}\right) + O\left(n^{-\frac{\zeta}{2q+1}}\right).$$

In addition,

$$\Delta_{i1}^{01} = |\epsilon_{i\alpha_1} - d_{ni\alpha_1} - (d_{ni\alpha_2} - \epsilon_{i\alpha_2})| = O\left(n^{-\frac{\zeta}{2q+1}}\right),$$

and

$$\Delta_{i1}^{10} = O\left(n^{-\frac{\zeta}{2q+1}}\right).$$

As such,

$$\Delta_{i1} \leq O\left(n^{\frac{-q-\theta}{2q+1}}\right) + O\left(n^{-\frac{\zeta}{2q+1}}\right).$$

Now consider

$$\begin{aligned}
\Delta_{i2} &\equiv \left| \frac{1}{2} n^{-\frac{2q}{2q+1}} f_{\alpha_2}(0|\boldsymbol{X}_i) K^d(0) \delta_2^2 - \frac{1}{2} n^{-\frac{2q}{2q+1}} f_{\alpha_1}(0|\boldsymbol{X}_i) K^d(0) \delta_1^2 \right| \\
&\leq \frac{1}{2} K^d(0) n^{-\frac{2q}{2q+1}} \left| f_{\alpha_2}(0|\boldsymbol{X}_i)\delta_2^2 - f_{\alpha_1}(0|\boldsymbol{X}_i)\delta_1^2 \right| \\
&= \frac{1}{2} K^d(0) n^{-\frac{2q}{2q+1}} \left| f_{\alpha_2}(0|\boldsymbol{X}_i)\left(\delta_2^2 - \delta_1^2\right) + f_{\alpha_2}(0|\boldsymbol{X}_i)\delta_1^2 - f_{\alpha_1}(0|\boldsymbol{X}_i)\delta_1^2 \right| \\
&= O\left(n^{\frac{-2q-\theta}{2q+1}} \sqrt{\log\log n}\right) + O\left(n^{\frac{-2q-\zeta}{2q+1}} \log\log n\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\Delta_i &\leq \Delta_{i1} + \Delta_{i2} \\
&\leq O\left(n^{\frac{-q-\theta}{2q+1}}\right) + O\left(n^{-\frac{\zeta}{2q+1}}\right) + O\left(n^{\frac{-2q-\theta}{2q+1}} \sqrt{\log\log n}\right) + O\left(n^{\frac{-2q-\zeta}{2q+1}} \log\log n\right) \\
&= O\left(n^{\frac{-q-\theta}{2q+1}}\right) + O\left(n^{-\frac{\zeta}{2q+1}}\right) \\
&\leq O\left(n^{-1-\eta}\right)
\end{aligned}$$

for θ, ζ sufficiently large. In this case,

$$\Delta \leq \sum_{i=1}^{n} \Delta_i = O\left(n^{-\eta}\right).$$

Now

$$\begin{aligned}
& \left| E\left[r_{ni}(\delta_1, \alpha_1) \right| \boldsymbol{X}_i] - E\left[r_{ni}(\delta_2, \alpha_2) \right| \boldsymbol{X}_i] \right| \\
\leq & \left| -\frac{1}{2} n^{-\frac{2q}{2q+1}} K^d(0) f_{\alpha_1}(0|\boldsymbol{X}_i)\delta_1^2 + \frac{1}{2} n^{-\frac{2q}{2q+1}} K^d(0) f_{\alpha_2}(0|\boldsymbol{X}_i)\delta_2^2 \right| \\
& +2 \left| \int_0^{|d_{ni\alpha_1}|} (u - |d_{ni\alpha_1}|) f_{\alpha_1}(u|\boldsymbol{X}_i) du - \int_0^{|d_{ni\alpha_2}|} (u - |d_{ni\alpha_2}|) f_{\alpha_2}(u|\boldsymbol{X}_i) du \right| \\
= & O\left(n^{\frac{-2q-\theta}{2q+1}} \sqrt{\log\log n} \right) + O\left(n^{\frac{-2q-\zeta}{2q+1}} \log\log n \right) + O\left(n^{-\frac{\zeta}{2q+1}} \right) \\
= & O\left(n^{\frac{-2q-\theta}{2q+1}} \sqrt{\log\log n} \right) + O\left(n^{-\frac{\zeta}{2q+1}} \right) \\
= & o\left(n^{\frac{-q-\theta}{2q+1}} \right) + O\left(n^{-\frac{\zeta}{2q+1}} \right) \\
\leq & O\left(n^{-1-\eta} \right)
\end{aligned}$$

for θ and ζ sufficiently large, which in turn implies that

$$\begin{aligned}
& \left| E\left[r_n(\delta_1, \alpha_1) \right| \mathcal{X}_n] - E\left[r_n(\delta_2, \alpha_2) \right| \mathcal{X}_n] \right| \\
\leq & \sum_{i=1}^{n} \left| E\left[r_{ni}(\delta_1, \alpha_1) \right| \boldsymbol{X}_i] - E\left[r_{ni}(\delta_2, \alpha_2) \right| \boldsymbol{X}_i] \right| \\
= & O\left(n^{-\eta} \right) \\
= & o(1)
\end{aligned}$$

for θ and ζ sufficiently large.

As such, Δ can be made to satisfy $\Delta = O\left(n^{-\eta}\right) = o(1)$ uniformly in probability for $\{\alpha_1, \alpha_2\} \subset [\tau, 1-\tau]$ with $|\alpha_1 - \alpha_2| \leq n^{-\frac{\zeta}{2q+1}}$ and $\{\delta_1, \delta_2\} \subset \{\delta : |\delta| \leq C\sqrt{\log\log n}\}$ with $|\delta_1 - \delta_2| \leq n^{-\frac{\theta}{2q+1}}$.

For an arbitrary $\xi > 0$, set $\eta \equiv (\xi + 1)\frac{\zeta+\theta}{2q+1}$. From (34) we have

$$\begin{aligned}
P\left[\sup_S |r_n(\delta, \alpha)| \geq (\xi+1) \frac{\zeta+\theta}{2q+1} \right] & \leq n^{-\frac{(\zeta+\theta)\xi}{2q+1}} \sqrt{\log\log n} \\
& \leq n^{-\xi} \sqrt{\log\log n} \\
& \to 0.
\end{aligned}$$

It follows that

$$P\left[\sup\left\{ |r_n(\delta, \alpha)| : |\delta| \leq C\sqrt{\log\log n},\ \alpha \in [\tau, 1-\tau] \right\} \geq (\xi+1) \frac{\zeta+\theta}{2q+1} \right] \to 0.$$

E.2 Proof of Lemma 2

Let

$$Z_{n,\alpha}(\boldsymbol{X}_i) \equiv n^{-\frac{q}{2q+1}} \sigma_\alpha(\boldsymbol{X}_i) \left(\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) \right) \psi_\alpha(\epsilon_{i\alpha}).$$

Note that

$$\begin{aligned} Var\left[Z_{n,\alpha}(\boldsymbol{X}_i) \middle| \boldsymbol{X}_i \right] &= \left(\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{m}_{n,\alpha}(\boldsymbol{X}_i) \right)^2 \alpha(1-\alpha) \\ &= O_p\left(n^{-\frac{2q}{2q+1}} \right), \end{aligned}$$

from the first conclusion of Proposition 1 and the assumption that the restricted estimators have the same pointwise rate of convergence under the null as their unrestricted counterparts. Note also that $S_{n,\alpha}$ is divergent, with

$$\begin{aligned} S_{n,\alpha} &= \left\{ \alpha(1-\alpha) \sum_{i=1}^{n} \left(\hat{m}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{m}_{n,\alpha}(\boldsymbol{X}_i) \right)^2 \right\}^{\frac{1}{2}} \\ &= n^{-\frac{q}{2q+1}} \left\{ \alpha(1-\alpha) \sum_{i=1}^{n} \sigma_\alpha^2(\boldsymbol{X}_i) \left(\hat{\delta}_{n,\alpha}(\boldsymbol{X}_i) - \tilde{\delta}_{n,\alpha}(\boldsymbol{X}_i) \right)^2 \right\}^{\frac{1}{2}} \\ &= O_p\left(n^{\frac{1}{4q+2}} \right). \end{aligned}$$

We have $E\left[\left| Z_{n,\alpha}(\boldsymbol{X}_i) \right|^3 \middle| \boldsymbol{X}_i \right] = O_p\left(n^{-\frac{3q}{2q+1}} \right)$, and therefore

$$\sum_{i=1}^{n} \left\{ \frac{E\left[\left| Z_{n,\alpha}(\boldsymbol{X}_i) \right|^3 \middle| \boldsymbol{X}_i \right]}{S_{n,\alpha}^3} \right\} \xrightarrow{p} 0.$$

It follows from Liapounov's theorem[20] that

$$\frac{1}{S_{n,\alpha}} \sum_{i=1}^{n} Z_{n,\alpha}(\boldsymbol{X}_i) \Bigg| \mathcal{X}_n \xrightarrow{d} N(0,1).$$

The desired conclusion follows.

References

Abadie, A. (2002) 'Bootstrap tests for distributional treatment effects in instrumental variable models.' *Journal of the American Statistical Association* 97, 284–292

Andrews, D. W. K. (1993) 'Tests for parameter instability and structural change with unknown change point.' *Econometrica* 61, 821–856

[20]Cf. e.g., Chung (2001, Theorem 7.1.2).

Billingsley, P. (1999) *Convergence of Probability Measures,* second ed. (New York: John Wiley & Sons)

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006) 'What mean impacts miss: Distributional effects of welfare reform experiments.' *American Economic Review* 96(4), 988–1012

Chernozhukov, V., and I. Fernández-Val (2005) 'Subsampling inference on quantile regression processes.' *Sankhyā* 67, 253–276

Chung, K. L. (2001) *A Course in Probability Theory,* third ed. (San Diego: Academic Press)

de Gooijer, J. G., and D. Zerom (2003) 'On additive conditional quantiles with high-dimensional covariates.' *Journal of the American Statistical Association* 98, 135–146

De Long, D. M. (1981) 'Crossing probabilities for a square root boundary by a Bessel process.' *Communications in Statistics, Part A—Theory and Methods* 10, 2197–2213

Doksum, K., and J.-Y. Koo (2000) 'On spline estimators and prediction intervals in nonparametric regression.' *Computational Statistics and Data Analysis* 35, 76–82

Durbin, J. (1973) *Distribution Theory for Tests Based on the Sample Distribution Function* (Philadelphia: Society for Industrial and Applied Mathematics)

Fan, J., and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications* (Boca Raton, Florida: Chapman & Hall/CRC)

Fan, J., and J. Jiang (2005) 'Nonparametric inferences for additive models.' *Journal of the American Statistical Association* 100, 890–907

Fan, J., C. Zhang, and J. Zhang (2001) 'Generalized likelihood ratio statistics and Wilks phenomenon.' *Annals of Statistics* 29, 153–193

Guerre, E., and P. Lavergne (2002) 'Optimal minimax rates for nonparametric specification testing in regression models.' *Econometric Theory* 18, 1139–1171

Hall, P., and S. J. Sheather (1988) 'On the distribution of a Studentized quantile.' *Journal of the Royal Statistical Society, Series B (Methodological)* 50, 381–391

Hastie, T. J., and R. J. Tibshirani (1990) *Generalized Additive Models* (London: Chapman and Hall)

Hawkins, D. L. (1987) 'A test for a change point in a parametric model based on a maximal Wald-type statistic.' *Sankhyā* 49, 368–376

Heckman, J. J., J. Smith, and N. Clements (1997) 'Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts.' *Review of Economic Studies* 64, 487–535

Horowitz, J. L., and S. Lee (2002) 'Semiparametric methods in applied econometrics: Do the models fit the data?' *Statistical Modelling* 2, 3–22

— (2005) 'Nonparametric estimation of an additive quantile regression model.' *Journal of the American Statistical Association* 100, 1238–1249

Ingster, Y. I. (1982) 'Minimax nonparametric detection of signals in white Gaussian noise.' *Problems of Information Transmission* 18, 130–140

— (1993a) 'Asymptotically minimax hypothesis testing for nonparametric alternatives, I.' *Mathematical Methods of Statistics* 2, 85–114

— (1993b) 'Asymptotically minimax hypothesis testing for nonparametric alternatives, II.' *Mathematical Methods of Statistics* 2, 171–189

— (1993c) 'Asymptotically minimax hypothesis testing for nonparametric alternatives, III.' *Mathematical Methods of Statistics* 2, 249–268

Khmaladze, E. V. (1981) 'Martingale approach in the theory of goodness-of-fit tests.' *Theory of Probability and its Applications* 26, 240–257

— (1988) 'An innovation approach to goodness-of-fit tests in $\mathbb{R}^m$.' *Annals of Statistics* 16, 1503–1516

Kiefer, J. (1959) 'K-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests.' *Annals of Mathematical Statistics* 30, 420–447

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

Koenker, R., and Z. Xiao (2002) 'Inference on the quantile regression process.' *Econometrica* 70, 1583–1612

Shorack, G. R. (1979) 'The weighted empirical process of row independent random variables with arbitrary distribution functions.' *Statistica Neerlandica* 33, 169–189

Spokoiny, V. G. (1996) 'Adaptive hypothesis testing using wavelets.' *Annals of Statistics* 24, 2477–2498

Table 1: Size and local power performance of a 5%-GLR test for location shift

	c					
	0	.2	.4	.6	.8	1.0
Rejection frequencies:	.07	.25	.31	.35	.59	.75

Notes:

1. 100 simulated random samples of size $n = 100$ were drawn from the model

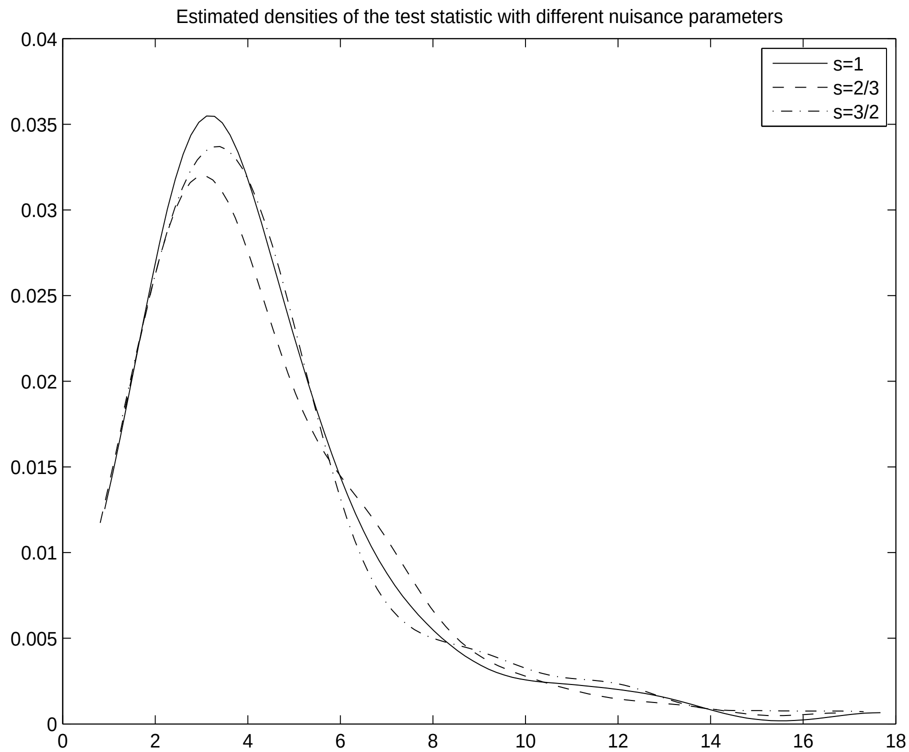
$$Y_i = .75X_{i1} + 1.5 \sin\left(.5\pi X_{i2}^2\right) + \left[1 + c\left(X_{i2} + .25X_{i2}^2\right)\right]\epsilon_i,$$

   where $\epsilon_i \sim N(0,1)$ and $(X_{i1}, X_{i2})^T$ is bivariate normal and independent of ϵ_i with mean zero, unit variance and covariance .2, and where $c \in \{0, .2, .4, .6, .8, 1.0\}$.

2. The null hypothesis of interest is of X_{i2} exerting a pure location-shift effect on the conditional distribution of Y_i. This hypothesis is stated more formally in (18).

3. The GLR test statistic $\sup_{\alpha\in[.15,.85]}\left(\frac{\lambda_n(H_0,\alpha)-b_{n,\alpha}(K)}{S_{n,\alpha}}\right)^2$ was approximated with $\max_{\alpha_g\in\hat{\mathcal{A}}}\left(\frac{\lambda_n(H_0,\alpha_g)-\hat{b}_{n,\alpha_g}(K)}{\hat{S}_{n,\alpha_g}}\right)^2$, where $\hat{\mathcal{A}}$ is a grid of 100 equally-spaced points covering the interval $[.15, .85]$; and where $\hat{b}_{n,\alpha_g}(K)$ and $\hat{S}_{n,\alpha_g}$ were constructed according to the description in Section 5.1.

4. The 5% critical value of the limiting quantity $\sup_{\alpha\in[.15,.85]} Q^2(\alpha)$, where $Q(\alpha)$ is as given in (13), is 8.85 (Andrews, 1993, Table 1).



Notes:

1. 100 simulated random samples of size $n = 100$ were drawn from the model

$$Y_i = .75X_{i1} + 1.5s\sin\left(.5\pi X_{i2}^2\right) + .25\epsilon_i,$$

where $\epsilon_i \sim N(0,1)$ and $(X_{i1}, X_{i2})^T$ is bivariate normal and independent of ϵ_i with mean zero, unit variance and covariance .2, and where $s \in \left\{1, \frac{2}{3}, \frac{3}{2}\right\}$.

2. The null hypothesis of interest is of X_{i2} exerting a pure location-shift effect on the conditional distribution of Y_i. This hypothesis is stated more formally in (18).

3. The GLR test statistic $\sup_{\alpha\in[.15,.85]}\left(\frac{\lambda_n(H_0,\alpha)-b_{n,\alpha}(K)}{S_{n,\alpha}}\right)^2$ was approximated with $\max_{\alpha_g\in\hat{\mathcal{A}}}\left(\frac{\lambda_n(H_0,\alpha_g)-b_{n,\alpha_g}(K)}{S_{n,\alpha_g}}\right)^2$, where $\hat{\mathcal{A}}$ is a grid of 100 equally-spaced points covering the interval $[.15, .85]$.

4. Estimates of the densities of the normalized test statistics were computed for each setting of s using a normal density kernel and a rule-of-thumb bandwidth $h_{100} = 1.06\hat{\sigma}_{100}100^{-.2}$, where $\hat{\sigma}_{100}$ is the standard error in simulations of the normalized test statistic.