University of Toronto
Department of Economics

Specification Analysis of Structural Quantile Regression Models

By Juan Carlos Escanciano and Chuan Goh

November 19, 2010

Specification Analysis of Structural Quantile Regression Models

Juan Carlos Escanciano[*]   Chuan Goh[†]
Indiana University   University of Toronto

First draft: 14 December 2009
This version: 23 September 2010[‡]

[*]Department of Economics, Indiana University, 105 Wylie Hall, 100 S. Woodlawn Ave., Bloomington, IN, U.S.A., 47405-7104. `jescanci@indiana.edu`, `http://mypage.iu.edu/~jescanci`.

[†]Department of Economics, University of Toronto, Max Gluskin House, 150 St. George St., Toronto, ON, Canada, M5S 3G7. `goh@economics.utoronto.ca`, `http://www.chuangoh.org`.

Abstract

This paper introduces a broad family of tests for the hypothesis of linearity in parameters of functions that are identified by conditional quantile restrictions involving instrumental variables. These tests are tantamount to assessments of lack of fit for quantile regression models involving endogenous conditioning variables, and may be applied to assess the validity of post-estimation inferences regarding the counterfactual effect of endogenous treatments on the distribution of outcomes. We show that the use of an orthogonal projection on the tangent space of nuisance parameters at each quantile index improves power performance and facilitates the simulation of critical values via the application of simple multiplier-type bootstrap procedures. Monte Carlo evidence is included, along with an application to an empirical analysis of the structure of demand for a particular subsegment of the market for anti-bacterial drugs in India.

1 Introduction

Let Y be a random variable, and let $\boldsymbol{X}$ and $\boldsymbol{Z}$ be d- and k-dimensional random vectors, respectively, where $k \geq d$. Consider the continuum of conditional probability restrictions given by

$$P\left[Y \leq \boldsymbol{X}^\top \boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{Z}\right] = \alpha, \ \alpha \in (0, 1), \tag{1}$$

where $\boldsymbol{\beta}_0(\cdot)$ is measurable unknown function from $[0, 1]$ to a compact subset of $\mathbb{R}^d$. As such, (1) denotes a continuum of linear-in-parameters quantile-regression models in which the covariate vector $\boldsymbol{X}$ is possibly endogenous and $\boldsymbol{Z}$ is the corresponding vector of instruments. The identifiability of the interest parameter $\boldsymbol{\beta}_0(\cdot)$ in the context of (1) was shown by Chernozhukov and Hansen (2005) under a conditional rank invariance condition applied to $Y - \boldsymbol{X}^\top \boldsymbol{\beta}_0(\alpha)$ for each $\alpha \in (0, 1)$.[1] Estimators of $\boldsymbol{\beta}_0(\alpha)$ in the context of (1) have been developed by Amemiya (1982); Powell (1983); Chen and Portnoy (1996); Honoré and Hu (2004); Chernozhukov and Hansen (2006); Ma and Koenker (2006); Lee (2007) and Sakata (2007), amongst other authors. The "structural quantile-regression" model denoted by (1) has become increasingly popular in applied econometric analysis over the past decade. Recent applications can be found in the papers of Chernozhukov and Hansen (2004); Machado and Mata (2005); Forbes (2008) and Chernozhukov et al. (2009).

This paper develops tests for the linearity in parameters of the structural quantile function $\boldsymbol{X}^\top \boldsymbol{\beta}_0(\cdot)$ in (1) over $(0, 1)$.[2] As such, the hypothesis is that the continuum of conditional moment restrictions implied by (1) holds with probability one for some $\boldsymbol{\beta}_0(\cdot)$ in the corresponding parameter space, while the alternative is that there is at least one quantile $\alpha' \in (0, 1)$ such that for almost every $\boldsymbol{z}$ in the support of $\boldsymbol{Z}$, the relation given above in (1) does not hold for any vector $\boldsymbol{\beta}_0(\alpha') \in \mathbb{R}^d$. Such tests are important in applications because the conclusions of any post-estimation inferences based on estimates of $\boldsymbol{\beta}_0(\cdot)$ will be sensitive to the implicit assumption that the structural quantile function in (1) is linear in parameters for all quantiles $\alpha \in (0, 1)$.[3]

---

[1]In particular, for each quantile index $\alpha \in (0, 1)$, $Y - \boldsymbol{X}^\top \boldsymbol{\beta}_0(\alpha)$ needs to be at least distributionally invariant across different realizations of $\boldsymbol{Z}$.

[2]Although the analysis presented in this paper assumes that the hypothesis is that the relation given above in (1) holds for all quantiles $\alpha \in (0, 1)$, the same testing procedures derived below can be modified *mutatis mutandis* to accommodate a hypothesis to the effect that (1) holds only for all α in some proper subset of $(0, 1)$.

[3]The analysis of wage distributions via models of their conditional quantiles (e.g., Machado

While tests of the validity of a linear-in-parameters conditional quantile function against unspecified alternatives have already been developed in a number of different papers, the analysis to the best of our knowledge has to date been limited to a single quantile, generally taken without loss of generality to be the median.[4] The present paper extends and complements the existing literature by considering specification analysis for linearity in parameters over a continuum of quantiles.[5]

The specification tests proposed in this paper involve functionals of weighted empirical processes corresponding to the family of conditional moment restrictions implied by the relation given above in (1). A novel feature of our tests is the explicit acknowledgment of the fact that deviations from a hypothesized conditional-quantile model in the direction of the score cannot be distinguished from deviations that are still consistent with the null. In other words, testing procedures of the sort proposed in this paper inherently run the risk of incorrectly rejecting the null because of the existence under the null of a nuisance parameter $\boldsymbol{\beta}_0(\alpha)$ at each fixed quantile $\alpha \in (0,1)$. As such, it is always possible that deviations from a given model consistent with the null are caused by deviations within the parameter space of $\boldsymbol{\beta}_0(\alpha)$ for some fixed quantile α rather than by deviations that would properly lead to rejection of the null. This notion is incorporated in our proposed tests by adjusting the empirical-process weighting function to incorporate an orthogonal projection on the tangent space of nuisance parameters at each quantile $\alpha \in (0,1)$. The result is a test with improved power properties whose

---

and Mata, 2005) is an especially natural domain of application for the diagnostic tests proposed in this paper. In particular, wage distributions typically exhibit heaping or flat areas over certain ranges of quantiles—e.g., in ranges of quantiles in the immediate vicinity of that corresponding to a legislated minimum wage. These nonregular features of wage distributions make the fitting of linear models to wage quantiles in ranges of particular interest for policy purposes problematic. (We are grateful to Thomas Lemieux for pointing this out.)

[4]e.g., see the papers of Zheng (1998); Bierens and Ginther (2001); Horowitz and Spokoiny (2002); Whang (2006a) and Whang (2006b) for the case where $\boldsymbol{X}$ contains no endogenous covariates. Horowitz and Lee (2009) develop a specification test for the more general case where $\boldsymbol{X}$ is possibly endogenous and the single-quantile restriction holds conditional on a vector of instruments.

[5]To the best of our knowledge, the only proposal for specification tests of the functional form of a conditional quantile model over a continuum of quantiles is given in Escanciano and Velasco (2006). These authors considered tests of possibly nonlinear dynamic quantile models implemented using subsampling. The methodology in the present paper differs from that considered in Escanciano and Velasco (2006) in several respects, most notably in the focus on iid data in the present paper. By way of contrast, Escanciano and Velasco (2006) study the time-series setting and propose test statistics that are not functionals of the same sort of weighted empirical process used to deliver feasible test statistics in the present paper.

asymptotic distribution is also amenable to a simple multiplier-type bootstrap approximation. This feature greatly simplifies the derivation of critical values in applications and is notably not shared by tests that are also based on weighted empirical processes corresponding to the conditional moment restrictions implied by (1) but whose weights are not adjusted by the orthogonal projection technique that we develop below. Although the use of the orthogonal projection that we propose here is motivated by the desire to improve the power properties of the resulting tests, the end result also involves the attractive feature of a family of tests having critical values that are convenient to simulate in practice.

The remainder of this paper is organized as follows. In Section 2 we introduce the weighted empirical processes that constitute the basis upon which the new testing procedure for the continuum of conditional probability restrictions in (1) is developed. We study the asymptotic distribution of the proposed tests under the null as well as under fixed and local alternatives in Section 3. Section 4 discusses the use of a multiplier-bootstrap technique to approximate the asymptotic distributions of test statistics under the null as well as associated issues of implementation. Section 5 summarizes the results of Monte Carlo experiments designed to assess the finite-sample performance of our proposed testing procedures. Section 6 illustrates the applicability of the tests proposed here in the context of an empirical analysis of the structure of demand within a particular subsegment of the market for anti-bacterial drugs in India using data originally analyzed by Chaudhuri et al. (2006). Section 7 concludes. Proofs and detailed tables relevant to the empirical example are deferred to the appendix. Throughout this paper the symbol C is a generic constant that may change from one expression to another.

2 The Test Statistics and their Asymptotic Null Distribution

Let $\boldsymbol{W} \equiv (\boldsymbol{X}^\top, Y)^\top$ where $\boldsymbol{X}$ is d-variate, and let Θ denote a compact subset of $\mathbb{R}^d$. We consider testing the specification of the linear-in-parameters structural quantile model given above in (1). As such, the null hypothesis H_0 is given by

$$E\left[\psi_\alpha(\boldsymbol{W}, \boldsymbol{\beta}_0) | \boldsymbol{Z}\right] = 0$$

almost surely for some $\boldsymbol{\beta}_0 \in \mathcal{F}$ and for all $\alpha \in (0,1)$, where $\mathcal{F}$ is a family of uniformly bounded functions from $(0,1)$ to $\mathcal{F} \subset \mathbb{R}^d$, and where

$$\psi_\alpha(\boldsymbol{W}, \boldsymbol{\beta}_0) \equiv \alpha - 1\left\{Y - \boldsymbol{X}^\top \boldsymbol{\beta}_0(\alpha) \leq 0\right\}. \tag{2}$$

The alternative H_1 is that

$$P\left[E\left[\psi_{\alpha'}(\boldsymbol{W}, \boldsymbol{\beta}_0)|\, \boldsymbol{Z}\right]\right] > 0$$

for some $\alpha' \in (0, 1)$ and all $\boldsymbol{\beta}_0(\cdot) \in \mathcal{F}$.

Let $\mathcal{G}$ denote a class of measurable weighting functions. An implication of H_0 is the uncountable number of unconditional moment restrictions

$$E[\psi_{\alpha}(\boldsymbol{W}, \boldsymbol{\beta}_0)g(\boldsymbol{Z})] = 0 \tag{3}$$

for all $g(\cdot) \in \mathcal{G}$, some $\boldsymbol{\beta}_0(\cdot) \in \mathcal{F}$ and all $\alpha \in (0, 1)$. For consistency purposes, a relatively large class $\mathcal{G}$ is recommended. Recalling that the instrument vector $\boldsymbol{Z}$ is taken to be k-dimensional, where $k \geq d$, popular choices of weighting function are the indicator class $\mathcal{G} = \{\boldsymbol{z} \rightarrow 1(\boldsymbol{z} \leq \boldsymbol{u}) : \boldsymbol{u} \in \mathbb{R}^k\}$ (Stute (1997) or Andrews (1997)) or the class of exponential functions used by Escanciano and Velasco (2006). Further examples are discussed in Bierens and Ploberger (1997) and in Stinchcombe and White (1998). In a given sample, the event $\{\boldsymbol{Z} \leq \boldsymbol{z}\}$ is unlikely to occur for $\boldsymbol{z}$ lying in a large subset of $\mathbb{R}^k$, and as such we follow Escanciano (2007) in recommending the choice $g(\boldsymbol{Z}) = 1\left\{\boldsymbol{\gamma}^\top \boldsymbol{Z} \leq u\right\}$ for $\boldsymbol{\gamma} \in \mathbb{R}^k$ with $\|\boldsymbol{\gamma}\| = 1$ and $u \in \mathbb{R}$. Our theory covers all of these possible mentioned choices of weighting function as special cases and also applies to other classes of weighting function such as nonparametric families; see the discussion after the assumptions.

Given a random sample $\{(\boldsymbol{Z}_i^\top, \boldsymbol{X}_i^\top, Y_i)^\top\}_{i=1}^n$ of size n, it seems natural to construct test statistics based on the quantile error-weighted empirical process indexed by $g \in \mathcal{G}$ and $\alpha \in (0, 1)$, i.e., on

$$S_n(g, \alpha) \equiv n^{-1/2} \sum_{i=1}^n \psi_{\alpha}(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n)g(\boldsymbol{Z}_i), \tag{4}$$

where $\hat{\boldsymbol{\beta}}_n(\alpha)$ is a $\sqrt{n}$−consistent estimator of $\boldsymbol{\beta}_0(\alpha)$.[6] The null hypothesis is likely to hold when the process $S_n(g, \alpha)$ is "close" in an appropriate sense to zero for almost all $(g, \alpha) \in \mathcal{G} \times (0, 1)$. This approach was used by Escanciano and Velasco (2006) and has the appealing property of delivering consistent tests. It does not acknowledge, however, that $\boldsymbol{\beta}_0(\alpha)$ is a nuisance parameter in the testing procedure for each $\alpha \in (0, 1)$. The situation here is similar in spirit to the one addressed by Neyman (1959) in a fully parametric case. Since $\boldsymbol{\beta}_0(\cdot) \in \mathcal{F}$ is

---

[6]The IVQR estimator of Chernozhukov and Hansen (2006) is an obvious candidate for $\hat{\boldsymbol{\beta}}_n(\alpha)$ when the number of endogenous elements in $\boldsymbol{X}$ is small.

unknown, deviations from the assumed data-generating process under the null in the direction of the score function cannot be distinguished from deviations within the parametric model, i.e. from local deviations in $\boldsymbol{\beta}_0(\cdot)$ that are nevertheless consistent with the null. A simple way to incorporate this information in the test statistic is to construct a test that does not waste power in the direction of the score. More precisely, instead of the process $S_n(g, \alpha)$ we consider a feasible version of the "projected" quantile-weighted empirical process

$$
\begin{aligned}
&R_n(g, \alpha) \\
\equiv\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n) \\
& \cdot \left\{ g(\boldsymbol{Z}_i) - \boldsymbol{D}^\top(g, \hat{\boldsymbol{\beta}}_n(\alpha)) \boldsymbol{\Delta}^{-1}(\alpha) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \right\},
\end{aligned} \tag{5}
$$

where

$$
\boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}(\alpha)) \equiv f\left(\boldsymbol{X}_i^\top \boldsymbol{\beta}(\alpha) \middle| \boldsymbol{Z}_i\right) \boldsymbol{X}_i
$$

is a $d \times 1$ vector of scores,

$$
\boldsymbol{D}(g, \boldsymbol{\beta}(\alpha)) \equiv E\left[\boldsymbol{\delta}\left(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}(\alpha)\right) g(\boldsymbol{Z})\right]
$$

and

$$
\boldsymbol{\Delta}(\alpha) \equiv E\left[\boldsymbol{\delta}\left(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha)\right) \boldsymbol{\delta}^\top\left(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha)\right)\right].
$$

Unlike S_n, R_n is asymptotically free of the effect of working with an estimate $\hat{\boldsymbol{\beta}}_n$ of the nuisance parameter $\boldsymbol{\beta}_0$. In particular, under some regularity conditions,[7] we have for any compact subset $\mathcal{A}$ of $[0, 1]$ that

$$
\sup_{g \in \mathcal{G}} \sup_{\alpha \in \mathcal{A}} |R_n(g, \alpha) - R_{n0}(g, \alpha)| = o_p(1), \tag{6}
$$

where R_{n0} is defined as R_n but with $\boldsymbol{\beta}_0$ replacing $\hat{\boldsymbol{\beta}}_n$, i.e.,

$$
\begin{aligned}
&R_{n0}(g, \alpha) \\
\equiv\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \\
& \cdot \left\{ g(\boldsymbol{Z}_i) - \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha)) \boldsymbol{\Delta}^{-1}(\alpha) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \right\}.
\end{aligned} \tag{7}
$$

---

[7]In particular, see the statement of Theorem 1 below.

Property (6) is critical in the derivation of the simple bootstrap approximation developed later. It should be noted that the process R_n is not the only one satisfying this property; any empirical process of the form $n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_n(\alpha)) g(\boldsymbol{Z}_i)$ with $g(\cdot)$ orthogonal to the score $\boldsymbol{\delta}(\cdot, \cdot, \boldsymbol{\beta}_0)$ satisfies (6).[8]

A prominent example of a process that does not satisfy this orthogonality property but is nevertheless unaffected asymptotically by the effects of an estimated null nuisance parameter is that associated with the transformation of Khmaladze (1981).[9] The transformation of Khmaladze (1981) is essentially motivated by the goal of providing a test statistic with a limiting variance that is free of the effect of having to estimate a null nuisance parameter. The focus on R_n in the present paper is primarily motivated by power considerations, although the resulting test statistics have the attractive feature of being asymptotically distribution free and also of being amenable to the convenient multiplier bootstrap scheme discussed below in Section 4.

Our test statistics will be continuous functionals of the feasible analogue of the projected process given above in (5), namely,

$$
\begin{aligned}
&\hat{R}_n(g, \alpha) \\
\equiv\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n) \\
& \cdot \left\{ g(\boldsymbol{Z}_i) - \hat{\boldsymbol{D}}_n^\top(g, \hat{\boldsymbol{\beta}}_n(\alpha)) \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \right\},
\end{aligned} \tag{8}
$$

where

$$
\hat{\boldsymbol{D}}_n(g, \boldsymbol{\beta}(\alpha)) \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}(\alpha)) g(\boldsymbol{Z}_i);
$$

$$
\hat{\boldsymbol{\Delta}}_n(\alpha) \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\delta}}_n \left(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha) \right) \hat{\boldsymbol{\delta}}_n^\top \left(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha) \right);
$$

$$
\hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \equiv \hat{f}_{h_m} \left(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha) \middle| \boldsymbol{Z}_i \right) \boldsymbol{X}_i,
$$

and where

$$
\hat{f}_{h_m}(u \mid \boldsymbol{Z}_i) \equiv \frac{1}{m h_m} \sum_{j=1}^{m} K \left(\frac{u - \boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha_j)}{h_m} \right). \tag{9}
$$

---

[8]i.e., the requirement is that $E[g(\boldsymbol{Z}) \boldsymbol{\delta}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0)] = \boldsymbol{0}$ for all $\alpha \in (0, 1)$.

[9]In particular, see Koenker and Xiao (2002).

In the context of (9), $K(\cdot)$ denotes a smoothing kernel, $\{\alpha_j\}_{j=1}^m$ is a discrete grid of evenly spaced points in $(0,1)$ that becomes dense in $[0,1]$ as $m \to \infty$, and $\{h_m\}$ is a sequence of bandwidths converging to zero at a suitable rate as $m \to \infty$. The artificial sample size m will be made to depend on n in our theory. The conditional density estimator in (9), whose form is directly motivated by the restrictions imposed by the null hypothesis, has a critical advantage over the generic conditional density estimator (Rosenblatt, 1956; Parzen, 1962) in which the Rosenblatt-Parzen multivariate kernel density estimator appears in the numerator. In particular, the rate of convergence of $\hat{f}_{h_m}(\cdot \mid \boldsymbol{Z}_i)$ corresponds to that of the Rosenblatt-Parzen estimator with univariate regressors, regardless of the dimension of $\boldsymbol{Z}$. One can also consider other estimators of the conditional density in the theory given here provided, of course, that they satisfy a suitable uniform convergence property.

In the final analysis we propose a test statistic that is a continuous functional of $\hat{R}_n$. A prominent example of such a function involving an indicator weighting function is a statistic of the Cramer-von Mises (CvM) type, i.e.,

$$CvM_n \equiv \int_{\mathbb{R}^d \times \mathcal{A}} \left| \hat{R}_n(1(\cdot \leq \boldsymbol{z}), \alpha) \right|^2 dF_{n,\boldsymbol{Z}}(\boldsymbol{z}) d\alpha, \tag{10}$$

where $F_{n,\boldsymbol{Z}}$ is the empirical distribution function of $\{\boldsymbol{Z}_i\}_{i=1}^n$ and $\mathcal{A}$ denotes either $[0,1]$ or a compact subset of $(0,1)$ according to the researcher's interest in modelling certain ranges of conditional quantiles. Computation of the functional given above in (10) is discussed in Section 5. In particular, CvM_n is chosen over alternative functionals in the simulation experiments discussed below for its relative simplicity of computation.

Despite what might be suggested by the uniform convergence in (6), we shall prove that, as is the case with the process S_n given above in (4), the limiting distribution of functionals of $\hat{R}_n$ such as CvM_n in (10) above will still depend on the underlying data-generating process and the null nuisance parameter $\boldsymbol{\beta}_0(\cdot)$. The crucial difference between the limiting distribution of functionals of $\hat{R}_n$ and the limiting distribution of functionals of S_n is that the distribution of functionals of $\hat{R}_n$ will not depend in the limit on the estimator $\hat{\boldsymbol{\beta}}_n$. This feature suffices to admit the existence of the simple bootstrap approximation discussed below in Section 4.

2.1 Asymptotic null distribution

In what follows we establish the limiting distribution of the quantile-weighted empirical process $\hat{R}_n$ given above in (8) under the null hypothesis H_0. The lim-

9

iting null distributions of the tests are the limit distributions of continuous functionals of $\hat{R}_n$ under H_0. To derive asymptotic results we consider the following notation and assumptions. Throughout the paper the family $\mathcal{F}$ in which the parameter $\boldsymbol{\beta}_0(\cdot)$ is assumed to take values is endowed with the supremum norm, i.e., $\|\boldsymbol{\beta}\|_{\mathcal{F}} \equiv \sup_{\alpha \in (0,1)} |\boldsymbol{\beta}(\alpha)|$. We study the weak convergence of $\hat{R}_n$ and related processes as elements of $l^\infty(\Pi)$, the space of all real-valued functions that are uniformly bounded on Π, where $\Pi \equiv \mathcal{G} \times (0,1)$. The space $l^\infty(\Pi)$ is furnished with the supremum norm, say $\|\cdot\|_\infty$; let $\mathcal{B}_{d_\infty}$ denote the corresponding Borel σ-algebra. Let $\Rightarrow$ denote weak convergence on $(l^\infty(\Pi), \mathcal{B}_{d_\infty})$ in the sense of Hoffmann-Jørgensen.[10] Note that $\Rightarrow$ denotes weak convergence on compacta. Let $N_{[\cdot]}(\delta, \mathcal{H}, \|\cdot\|)$ denote the δ-bracketing number of a class of functions $\mathcal{H}$ with respect to a norm $\|\cdot\|$, i.e., the smallest number r such that there exist $f_1, ..., f_r$ and $\Delta_1, ..., \Delta_r$ with $\max_{1 \le i \le r} \|\Delta_i\| < \delta$ and $\|f - f_i\| < \Delta_i$ for some $i \in \{1, \ldots, r\}$ for all $f \in \mathcal{H}$.[11] For a class $\mathcal{H}$ of measurable functions define the envelope function $H(X)$ as a measurable function satisfying $|h(X)| < H(X)$ for all $h \in \mathcal{H}$. Throughout the proofs denote by $\|\cdot\|_{p,P}$ the L_p-norm with respect to P, i.e.,

$$\|f\|_{p,P} = \left(\int |f(x)|^p \, dP(x) \right)^{1/p}.$$

When P is the underlying probability measure we write $\|\cdot\|_p \equiv \|\cdot\|_{p,P}$ for simplicity. Furthermore, when $p = 2$, write $\|\cdot\| \equiv \|\cdot\|_2$. Henceforth, weak convergence and almost sure convergence of nonmeasurable maps is understood, as usual, in the sense of outer almost sure convergence.[12]

Regularity conditions underlying the derivation of the asymptotic distribution of our proposed test statistics under the null are given as follows.

Assumption 1 (Data-generating process). *1.*

$$\{\boldsymbol{W}_i\}_{i=1}^n = \left\{ (\boldsymbol{Z}_i^\top, \boldsymbol{X}_i^\top, Y_i)^\top : i = 1, \ldots, n \right\}$$

is a sequence of iid random $(k + d + 1)$-variates, where $k \ge d$.

2. For each realization $\boldsymbol{x}$ of $\boldsymbol{X}$,

$$Y = Y_{\boldsymbol{x}} = q(\boldsymbol{x}, U_{\boldsymbol{x}}),$$

where $U_{\boldsymbol{x}}$ is uniformly distributed on $(0,1)$, and and $q(\boldsymbol{x}, \alpha)$ is strictly increasing for all $\alpha \in (0,1)$.

---

[10]e.g., Dudley (1999, p. 94) or van der Vaart and Wellner (1996, Definition 1.3.3).

[11]See van der Vaart and Wellner (1996, Definition 2.1.6).

[12]See van der Vaart and Wellner (1996) for definitions. We omit a discussion here.

3. *For each realiation $\boldsymbol{x}$ of $\boldsymbol{X}$, $U_{\boldsymbol{x}}$ is independent of $\boldsymbol{Z}$.*

4. *For each pair of realizations $\boldsymbol{x}$ and $\boldsymbol{x}'$ of $\boldsymbol{X}$, we either have*

$$U_{\boldsymbol{x}} = U_{\boldsymbol{x}'}$$

   almost surely, or

$$U_{\boldsymbol{x}} \sim U_{\boldsymbol{x}'},$$

   where $\sim$ denotes distributional equivalence.

5. *For all $\delta \in (0, 1]$ and all $\alpha_1 \in (0, 1)$,*

$$\sup_{\alpha_2:|\alpha_1-\alpha_2|<\delta} \|\boldsymbol{\beta}_0(\alpha_1) - \boldsymbol{\beta}_0(\alpha_2)\|^2 < C\delta.$$

6. *The family of conditional distribution functions*

$$\left\{ F\left(u|\, \boldsymbol{z} \right) : \, \boldsymbol{x} \in \mathbb{R}^k \right\}$$

   has corresponding densities with respect to Lebesgue measure given by

$$\left\{ f\left(u|\, \boldsymbol{z} \right) : \, \boldsymbol{z} \in \mathbb{R}^k \right\}$$

   that are uniformly bounded with uniformly bounded derivatives to fourth order with respect to $u \in \mathbb{R}$.

7. *For all $\alpha \in (0, 1)$, $\boldsymbol{\Delta}(\alpha) \equiv E\left[\boldsymbol{\delta}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}(\alpha))\boldsymbol{\delta}^\top(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}(\alpha)) \right]$ is nonsingular in a neighbourhood of $\boldsymbol{\beta}(\alpha) = \boldsymbol{\beta}_0(\alpha)$.*

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

Assumption 2 (Estimator of the structural parameter under the null)**.** *The estimator $\hat{\boldsymbol{\beta}}_n(\cdot)$ of $\boldsymbol{\beta}_0(\cdot)$ satisfies the following under the restrictions of the null hypothesis:*

1. $\hat{\boldsymbol{\beta}}_n \in \mathcal{F}$ *with probability tending to one;*

2. $\left\| \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \right\|_{\mathcal{F}} = O_p(n^{-1/2}).$

Assumption 3 (Weighting functions). *The class of functions $\mathcal{G}$ has envelope G satisfying $\|G\|_p < \infty$ for some $p \geq 4$ and is endowed with a norm $\|\cdot\|_{\mathcal{G}}$ that satisfies for all $\delta \in (0, 1]$*

$$E\left[\sup_{\|g_1 - g_2\|_{\mathcal{G}}} |g_1 - g_2|^2\right] < C\delta$$

and

$$\int_0^\infty \left(\log N(\delta^2, \mathcal{G}, \|\cdot\|_{\mathcal{G}})\right)^{1/2} d\delta < \infty.$$

Assumption 4 (Kernel and bandwidth). *1. $K(u)$ satisfies the following conditions:*

    (a) *$K = \Psi_1 - \Psi_2$, where Ψ_1 and Ψ_2 are bounded, non-decreasing and right-continuous functions.*

    (b) *$\|K\|_\infty \equiv \sup_u |K(u)| = \kappa$ for some $\kappa \in (0, \infty)$.*

    (c) *$\int_{-\infty}^\infty K(u)du = 1$.*

    (d) *K satisfies a Lipschitz condition on $\mathbb{R}$.*

    (e) *K is of second order, i.e., $\int_{-\infty}^\infty uK(u) = 0$, $\int_{-\infty}^\infty u^2 K(u) = \mu_{2K}$ for some $\mu_{2K} \in (0, \infty)$ and $\int_{-\infty}^\infty [K(u)]^2 du = B$ for some $B \in (0, \infty)$.*

2. $h_m \in \left[c\left(\frac{\log m}{m}\right)^{1-\zeta}, c^{-1}\left(\frac{\log m}{m}\right)^{1-\zeta}\right]$ for some $c, \zeta \in (0, 1)$.

As was shown by Chernozhukov and Hansen (2005), the first three conditions of Assumption 1 are sufficient for the basic restriction

$$P\left[Y \leq q(\boldsymbol{X}, \alpha) | \boldsymbol{Z}\right] = \alpha$$

to hold with probability one for all $\alpha \in (0, 1)$, and enables the existence of estimators of structural parameters satisfying the conditions of Assumption 2. The conditions of Assumption 2 are notably satisfied by the IVQR procedure of Chernozhukov and Hansen (2006), although of course any estimator of the null nuisance parameter $\boldsymbol{\beta}_0(\cdot)$ can be incorporated in the test proposed below. Conditions sufficient for the entropy requirement of Assumption 3 can be found in van der Vaart and Wellner (1996). All popular parametric choices mentioned earlier for the weighting function class $\mathcal{G}$ satisfy this assumption. For example, the indicator family satisfies the requirements of Assumption 3 with $\|\cdot\|_{\mathcal{G}} = \|\cdot\|_2$ under a mild

continuity condition on the conditional distribution function of $\boldsymbol{X}$.[13] Other examples of function classes satisfying the conditions of Assumption 3 are spaces of smooth functions including those associated with the names of Sobolev, Hölder and Besov. For these classes, the covering number condition in Assumption 3 can be found in many books and articles on approximation theory.[14] Finally, we note that Assumption 4 allows for the use of the most popular smoothing kernels in empirical practice, including in particular the Gaussian kernel.

We are ready now to establish the asymptotic distribution of $\hat{R}_n$. The proof, which is given in Appendix A.2, proceeds in two steps. The first step involves showing that $\hat{R}_n$ is asymptotically equivalent under the null and $n^{-\frac{1}{2}}$-local alternatives to the process R_{n0} given above in (7). In the second step we analyze the weak convergence of the process R_{n0}.

Theorem 1. *Suppose the conditions of Assumptions 1–4 hold. Also assume the validity of the hypothesis*

$$E\left[\psi_\alpha(\boldsymbol{W},\boldsymbol{\beta}_0)|\,\boldsymbol{Z}\right] = 0$$

almost surely for some $\boldsymbol{\beta}_0 \in \mathcal{F}$ and for all $\alpha \in (0,1)$, where $\mathcal{F}$ denotes a family of uniformly bounded functions from $(0,1)$ to $\mathcal{F} \subset \mathbb{R}^d$, and where

$$\psi_\alpha(\boldsymbol{W},\boldsymbol{\beta}_0) \equiv \alpha - 1\left\{Y - \boldsymbol{X}^\top\boldsymbol{\beta}_0(\alpha) \leq 0\right\}.$$

Then for any compact set $\mathcal{A} \subset [0,1]$, the following convergence holds:

$$\sup_{g\in\mathcal{G}}\sup_{\alpha\in\mathcal{A}}\left|\hat{R}_n(g,\alpha) - R_{n0}(g,\alpha)\right| = o_p(1).$$

Proof. See Appendix A.2. □

---

[13]See van der Vaart and Wellner (1996, p. 85) for further details.

[14]To give an example, define for any vector $(a_1, \ldots, a_k)$ of k integers the differential operator $D^a \equiv \frac{\partial^{|a|}}{\partial z_1^{a_1}\cdots\partial z_q^{a_k}}$, where $|a| = \sum_{i=1}^k a_i$. Let R be a bounded, convex subset of $\mathbb{R}^k$ with a nonempty interior. For any smooth function $h : R \subset \mathbb{R}^k \to \mathbb{R}$ and some $\eta > 0$, let $\lfloor\eta\rfloor$ be the largest integer smaller than η, and define

$$\|h\|_{\infty,\eta} \equiv \max_{|a|\leq\lfloor\eta\rfloor}\sup_{\boldsymbol{x}}|D^a h(\boldsymbol{x})| + \max_{|a|=\lfloor\eta\rfloor}\sup_{\boldsymbol{x}_1\neq\boldsymbol{x}_2}\frac{|D^a h(\boldsymbol{x}_1) - D^a h(\boldsymbol{x}_2)|}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{\eta-\lfloor\eta\rfloor}}.$$

Furthermore, let $C_c^\eta(R)$ be the set of all continuous functions $h : R \subset \mathbb{R}^k \to [0,1]^k$ with $\|h\|_{\infty,\eta} \leq c$. If $\mathcal{G} = C_c^\eta(R)$, then $\mathcal{G}$ satisfies Assumption 2 with $\|\cdot\|_\mathcal{G} = \|\cdot\|_{\infty,\eta}$ provided that $\eta > 2k$; see van der Vaart and Wellner (1996, Theorem 2.7.1).

Theorem 1 indicates that $\hat{R}_n$ behaves like R_{n0} in large samples; in particular, the limiting behaviour of $\hat{R}_n$ may be approximated by that of R_{n0}, as indicated by the following result.

Corollary 1. *Under the conditions of Theorem 1 we have*

$$\hat{R}_n \Rightarrow R_\infty,$$

where R_∞ denotes a Gaussian process with mean zero and covariance function

$$V(g_1, g_2, \alpha_1, \alpha_2) \equiv (\min\{\alpha_1, \alpha_2\} - \alpha_1\alpha_2) E\left[g_1^\perp(\boldsymbol{Z}, \boldsymbol{X}, \alpha_1)g_2^\perp(\boldsymbol{Z}, \boldsymbol{X}, \alpha_2)\right],$$

where for $j = 1, 2$,

$$g_j^\perp(\boldsymbol{Z}, \boldsymbol{X}, \alpha_j) \equiv g_j(\boldsymbol{Z}) - \boldsymbol{D}^\top(g_j, \boldsymbol{\beta}_0(\alpha_j))\boldsymbol{\Delta}^{-1}(\alpha_j)\boldsymbol{\delta}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha_j)).$$

.

Proof. See Appendix A.3. □

It follows from Corollary 1, the continuous mapping theorem and the Glivenko-Cantelli Theorem that the asymptotic distribution of CvM_n is characterized by the convergence

$$CvM_n \Rightarrow CvM_\infty \equiv \int_{\mathbb{R}^d \times \mathcal{A}} |R_\infty(\mathbf{1}\{\cdot \leq \boldsymbol{z}\}, \alpha)|^2 \, dF_{\boldsymbol{Z}}(\boldsymbol{z})d\alpha.$$

Power properties of tests based on the process $\hat{R}_n$ are the subject of the next section of the paper.

3 Asymptotic Power Properties of the Proposed Test

In this section we study the consistency properties of tests based on continuous functionals $T(\hat{R}_n)$ of the process given in (8) above. We first focus on the asymptotic distribution of $\hat{R}_n$ under a certain sequence of local alternatives converging to null at a parametric rate $n^{-\frac{1}{2}}$. We consider the data-generating process for the sequence of local alternatives given by

$$H_{1,n}(a): E\left[\psi_\alpha\left(Y - \boldsymbol{X}^\top\boldsymbol{\beta}_0(\alpha)\right)\big| \boldsymbol{Z}\right] = \frac{a(\boldsymbol{Z}, \boldsymbol{X}, \alpha)}{\sqrt{n}} \tag{11}$$

14

almost surely for some $\beta_0 \in \mathcal{F}$ and all $\alpha \in \mathcal{A}$, where as before $\mathcal{A}$ denotes either the interval $[0, 1]$ or some compact subset of $(0, 1)$. We require the function $a(\cdot, \cdot, \alpha) : \mathbb{R}^{k+d} \to \mathbb{R}$ to satisfy the conditions of the following assumption. In particular, $a(\cdot, \cdot, \alpha)$ is required to be orthogonal to the score.[15]

Assumption 5 (Local alternatives). *The following holds for $\mathcal{A}$ either equal to $[0, 1]$ or to some compact subset of $(0, 1)$:*

1. *$a(\cdot, \cdot, \alpha)$ is such that*

$$E\left[\sup_{\alpha \in \mathcal{A}} |a(\boldsymbol{Z}, \boldsymbol{X}, \alpha)|\right] < \infty.$$

2. *There exists a random variable $b(\boldsymbol{Z}, \boldsymbol{X})$ with $E\left[b^2(\boldsymbol{Z}, \boldsymbol{X})\right] < \infty$, such that for all $\alpha_1, \alpha_2 \in \mathcal{A}$,*

$$|a(\boldsymbol{Z}, \boldsymbol{X}, \alpha_1) - a(\boldsymbol{Z}, \boldsymbol{X}, \alpha_2)| \leq b(\boldsymbol{Z}, \boldsymbol{X}) |\alpha_1 - \alpha_2|$$

almost surely.

3. *For all $\alpha \in \mathcal{A}$ we have*

$$E\left[\delta(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha))a(\boldsymbol{Z}, \boldsymbol{X}, \alpha)\right] = \boldsymbol{0}. \tag{12}$$

The "non-centrality parameter" under sequences of local alternatives given by $H_{1,n}(a)$ in (11) above takes the form

$$D_a(g, \alpha) \equiv E\left[a(\boldsymbol{Z}, \boldsymbol{X}, \alpha)g(\boldsymbol{Z})\right], \tag{13}$$

as is indicated by the following result.

Theorem 2. *Under the conditions of Assumptions 1–5 and under the sequence of local alternatives given by (11), we have*

$$\hat{R}_n \Rightarrow R_\infty + D_a,$$

where R_∞ is as given above in the statement of Corollary 1, and where D_a is as given above in (13).

---

[15]This does not entail any loss of generality. See in particular Escanciano (2009) and the remarks at the end of this section.

Proof. See Appendix A.4. □

Several conclusions emerge from the convergence in Theorem 2. First, our tests are able to detect *any* Pitman local alternative satisfying the orthogonality condition in (12). Second, suppose instead of $\hat{R}_n$ one entertains a classical test based on the weighted process S_n given in (4) above. Under the standard assumption that $\hat{\boldsymbol{\beta}}_n(\alpha)$ is asymptotically linear under $H_{1,n}(a)$ for all $\alpha \in \mathcal{A}$, i.e.,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)\right) = \boldsymbol{\xi}(a) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Lambda}^{-1}(\alpha)\boldsymbol{k}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0)\psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) + o_p(1),$$

where

$$\begin{aligned}
\boldsymbol{\xi}(a) &\equiv E\left[\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha))g(\boldsymbol{Z})\right], \\
\boldsymbol{\Lambda}(\alpha) &\equiv E\left[\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha))\boldsymbol{\delta}^\top(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha))\right]
\end{aligned}$$

and $\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha))$ is a measurable $(d \times 1)$-valued function, it can be similarly proved that under the same sequence of local alternatives $H_{1,n}(a)$ that

$$S_n(g, \cdot) \Rightarrow S_\infty(g, \cdot) + D_a. \tag{14}$$

Here $S_\infty(g, \alpha)$ is a zero-mean Gaussian process with covariance function

$$\tilde{V}(\tilde{g}_1, \tilde{g}_2, \alpha_1, \alpha_2) \equiv (\alpha_1 \wedge \alpha_2 - \alpha_1\alpha_2)E\left[\tilde{g}_1(\alpha_1)\tilde{g}_2(\alpha_2)\right],$$

where for $j = 1, 2$,

$$\tilde{g}_j(\alpha_j) \equiv g_j(\boldsymbol{Z}) - \boldsymbol{D}^\top(g_j, \boldsymbol{\beta}_0(\alpha_j))\boldsymbol{\Lambda}^{-1}(\alpha_j)\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha_j)).$$

Note that the same shift function D_a appears in both the statement of Theorem 2 and (14). It follows from standard properties of orthogonal projections that for all $g \in \mathcal{G}$ and all $\alpha \in \mathcal{A}$,

$$V(g, g, \alpha, \alpha) \leq \tilde{V}(g, g, \alpha, \alpha),$$

where $V(g, g, \alpha, \alpha)$ is as given above in the statement of Corollary 1. In other words, the process $R_\infty(g, \cdot)$ is of "smaller" magnitude than $S_\infty(g, \cdot)$. As a consequence of this, one can prove that tests based on $\hat{R}_n(g, \cdot)$ will have better power properties than tests based on $S_n(g, \cdot)$.

16

With respect to the global power properties of tests based on $\hat{R}_n$, let

$$g^\perp(\boldsymbol{Z}, \boldsymbol{X}, \alpha) \equiv g(\boldsymbol{Z}) - \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha))\boldsymbol{\Delta}^{-1}(\alpha)\boldsymbol{\delta}(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0(\alpha)).$$

We have from the proof of Theorem 1 that under any alternative,

$$n^{-\frac{1}{2}}\hat{R}_n(g, \alpha) \Rightarrow E\left[g^\perp(\boldsymbol{Z}, \boldsymbol{X}, \alpha)\psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0)\right] \tag{15}$$

uniformly for $g \in \mathcal{G}$ and $\alpha \in \mathcal{A}$. It follows that our proposed test is consistent against all alternatives not collinear with the score, that is, against alternatives characterized by measurable functions m such that

$$E\left[g^\perp(\boldsymbol{Z}, \boldsymbol{X}, \alpha)m\left(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\beta}_0\right)\big|\, \boldsymbol{Z}\right] \neq 0$$

with probability one over a set with positive $F\left(\cdot\,|\,\boldsymbol{Z}\right)$-measure. Note that this is not an important limitation—in particular, all tests based on integrated conditional moment restrictions have trivial local power against those directions.[16] As a result of this, the global power of all tests in the direction of the score will be also low.[17]

4   Bootstrap Approximation to the Asymptotic Distribution of the Test

The foregoing sections of this paper have shown that the asymptotic null distribution of continuous functionals of $\hat{R}_n$ is liable to depend in a complex way on the underlying data-generating process and the correctness of the model specification under the null hypothesis. It follows from this that critical values for test statistics based on continuous functionals of $\hat{R}_n$ cannot in general be tabulated for more than a few special cases. In this section we overcome this problem with the assistance of a multiplier-type bootstrap.

The literature on inference in the context of conditional-quantile models contains many proposals involving resampling.[18] The methods described in these proposals all require the computation of the parameter estimates $\hat{\boldsymbol{\beta}}_n(\alpha)$ at each

---

[16]See Escanciano (2009).

[17]See Strasser (1985).

[18]e.g., Hahn (1995), Horowitz (1998), Bilias et al. (2000), Sakov and Bickel (2000) or He and Hu (2002), among others.

bootstrap replication. Approaches based on subsampling[19] have also proven popular.[20] Approaches based on subsampling have the disadvantage, however, of relying crucially on the subjective choice of subsample size, which is known to have occasionally dramatic effects on the outcome of any resulting inferences. In particular, two researchers using the same data and working with the same model can reach different conclusions simply by virtue of having chosen different tuning parameters. This undesirable property is shared by a number of other inference methodologies, including the block bootstrap.

In this paper we consider a multiplier-type bootstrap that avoids the two disadvantages of existing methods cited above. In particular, the bootstrap method proposed here avoids the need to compute parameter estimates at each bootstrap replication. The new approach also does not involve the need to select tuning parameters and appears to be superior to other competing methods in terms of the finite-sample accuracy of the distributional approximation.[21]

The method advocated in this section involves approximating the asymptotic distribution of a smooth functional $\phi\left(\hat{R}_n\right)$ with that of $\phi\left(\hat{R}_n^*\right)$, where $\hat{R}_n^*$ is a simple multiplier-bootstrap approximation of $\hat{R}_n$ given by

$$
\begin{aligned}
&\hat{R}_n^*(g, \alpha) \\
\equiv\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \\
& \cdot \left\{ g(\boldsymbol{Z}_i) - \hat{\boldsymbol{D}}_n^\top(g, \hat{\boldsymbol{\beta}}_n(\alpha)) \hat{\Delta}_n^{-1}(\alpha) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \right\} V_i,
\end{aligned}
$$

where $\{V_i\}$ is a sequence of iid random variables with zero mean, unit variance, bounded support and also independent of the sequence $\{\boldsymbol{W}_i\}$. An example of a possible multiplier sequence $\{V_i\}$ involves iid Bernoulli variates with

$$
P\left[V_i = \frac{1}{2}\left(1 - \sqrt{5}\right)\right] = b; \tag{16}
$$

$$
P\left[V_i = \frac{1}{2}\left(1 + \sqrt{5}\right)\right] = 1 - b, \tag{17}
$$

---

[19] i.e., resampling fewer observations than exist in the original sample without replacement.

[20] e.g., Chernozhukov and Fernández-Val (2005), Whang (2006a) or Escanciano and Velasco (2006), among others.

[21] Section 5 presents some Monte Carlo evidence in support of this assertion.

where $b = \frac{1+\sqrt{5}}{2\sqrt{5}}$. Another example also involves iid Bernoulli variates, but with

$$
\begin{aligned}
P\left[V_i = 1\right] &= \frac{1}{2}; \\
P\left[V_i = -1\right] &= \frac{1}{2};
\end{aligned}
$$

see Wu (1986). The theoretical justification of this bootstrap approximation does not require any assumptions in addition to those already given above.

The consistency of the proposed bootstrap method involves the concept of convergence in distribution with probability one.[22] The unknown limiting distribution of $\phi\left(\hat{R}_n\right)$ under the null, i.e., the distribution of $\phi\left(R_\infty\right)$, is approximated by the bootstrap distribution of $\phi\left(\hat{R}_n^*\right)$. In other words, the bootstrap empirical distribution

$$
\hat{F}_n^*\left(x \middle| \{(\boldsymbol{Z}_i^\top, \boldsymbol{X}_i^\top, Y_i)\}_{i=1}^n\right) = P\left[\phi\left(\hat{R}_n^*\right) \le x \middle| \{(\boldsymbol{Z}_i^\top, \boldsymbol{X}_i^\top, Y_i)\}_{i=1}^n\right]
$$

is taken to be a consistent estimate of the asymptotic null distribution function

$$
F_\infty\left(x\right) = P\left[\phi\left(R_\infty\right) \le x\right].
$$

In this case, the null hypothesis will be rejected at the τ-level of significance when

$$
\phi\left(\hat{R}_n\right) \ge c_{n,\tau}^*,
$$

where $c_{n,\tau}^*$ is such that

$$
\hat{F}_n^*\left(c_{n,\tau}^* \middle| \{\boldsymbol{W}_i\}_{i=1}^n\right) = 1 - \tau.
$$

It is also possible to use bootstrap p-values in this context. For example, the null could be rejected whenever $p_n^* < \tau$, where

$$
p_n^* \equiv P\left[\phi\left(\hat{R}_n^*\right) \ge \phi\left(\hat{R}_n\right) \middle| \{(\boldsymbol{Z}_i^\top, \boldsymbol{X}_i^\top, Y_i)\}_{i=1}^n\right].
$$

This bootstrap-based test is clearly valid if $\hat{F}_n^*$ is a consistent estimator of F_∞ at each continuity point of F_∞. In the case of almost-sure consistency, an equivalent condition for validity is that $\phi\left(\hat{R}_n^*\right) \xrightarrow{d} \phi\left(R_\infty\right)$ almost surely.[23]

---

[22]A less restrictive concept is convergence in distribution in probability; see Giné and Zinn (1990).

[23]Further details are available in Giné and Zinn (1991) or van der Vaart and Wellner (1996).

Theorem 3. *Suppose the conditions of Assumptions 1 and 3 hold. Then for any continuous functional* $\phi(\cdot)$,

$$\phi\left(\hat{R}_n^*\right) \xrightarrow{d} \phi\left(R_\infty\right)$$

almost surely.

Proof. See Appendix A.5. □

It is straightforward to show that Theorem 3 implies the consistency of our multiplier-bootstrap test against all alternatives not collinear to the score, provided that ϕ is such that $\phi(f) = 0 \Leftrightarrow f = 0$ a.s.-$P\left[\cdot\,|\,\boldsymbol{Z}\right]$. Moreover, it can be proved that our bootstrap-based test preserves the asymptotic local power properties of $\phi\left(\hat{R}_n\right)$, including its asymptotic admissibility. Details have been omitted in order to economize on space.

5   Numerical Evidence

This section presents the result of a Monte Carlo experiment designed to evaluate the finite-sample performance of our proposed tests relative to more immediately familiar tests based on subsampling. For simplicity, the focus in this section is on the "non-structural" special case of the structural quantile model where the covariate vector $\boldsymbol{X}$ does not contain any endogenous components, and where as such the instrument vector satisfies $\boldsymbol{Z} = \boldsymbol{X}$.

In this connection, we considered two data-generating processes for our simulations. The first ($DGP1$) is given by

$$DGP1 : Y_i = X_{1i} + X_{2i} + c\sigma_i^{\frac{3}{2}} + u_i, \ i = 1, \ldots, n; \tag{18}$$

while the second ($DGP2$) is given by

$$DGP2 : Y_i = X_{1i} + X_{2i} + \left(1 + c\sigma_i^{\frac{3}{2}}\right) u_i, \ i = 1, \ldots, n; \tag{19}$$

where

$$\sigma_i \equiv X_{1i}^2 + X_{2i}^2 + X_{1i}X_{2i}.$$

X_{1i}, X_{2i} and u_i are taken to be iid $N(0, 1)$ and mutually independent. Let

$$\boldsymbol{X}_i \equiv \left(X_{1i}, X_{2i}\right)^\top.$$

In the context of $DGP1$ and $DGP2$, the hypothesis given above in (1) corresponds to the location-shift model with $c = 0$, so the model for the conditional quantile function under the null is simply the quantile-regression model given by

$$F_{Y_i|\boldsymbol{X}_i}^{-1}(\alpha) = \tilde{\boldsymbol{X}}_i^\top \boldsymbol{\beta}_0(\alpha)$$

for each $\alpha \in (0, 1)$, where $\tilde{\boldsymbol{X}}_i = (\ 1 \quad X_{1i} \quad X_{2i}\ )^\top$ and $\boldsymbol{\beta}_0(\alpha) = (\Phi^{-1}(\alpha), 1, 1)^\top$. Here we let $\Phi^{-1}(\alpha)$ denote the quantile function of a standard normal random variable. Note that for this model the score vector satisfies

$$\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{X}_i, \boldsymbol{\beta}(\alpha)) = \phi\left(\Phi^{-1}(\alpha)\right)\boldsymbol{X}_i,$$

where $\phi(\cdot)$ is the standard normal density function. Throughout the simulations the estimator $\hat{\boldsymbol{\beta}}_n(\alpha)$ of the null parameter vector will be taken to be the regression α-quantile of Koenker and Bassett (1978).[24]

We considered two sample sizes $n = 100$ and $n = 300$ and a subinterval of quantiles given by $\mathcal{A} = [0.1, 0.9]$. The number of Monte Carlo replications was set to 1000. In order to economize on space we only report results for simulations with $n = 100$, as the results for $n = 300$ were qualitatively similar. We consider an approximation of the Cramér-von Mises test defined in (10) over a grid of $m = 30$ evenly spaced points in the interval $[\epsilon, 1 - \epsilon] \equiv [0.1, 0.9]$. Denote by $\{\alpha_j\}_{j=1}^m$ the points in the grid, with $\epsilon = \alpha_1 < \cdots < \alpha_m = 1 - \epsilon$. Two versions of the Cramér-von Mises test were considered. One version is infeasible in the sense that the score vector $\boldsymbol{\delta}$ is assumed to be known, which in the case of the model under consideration corresponds to the assumption that the researcher has knowledge both of the existence of the location-shift effect of the covariates on the distribution of the outcome variable and of the actual distribution of the latent disturbances in the model. The second version of the CvM test considered here is empirically feasible and involves the preliminary estimate of the conditional density $f(u|\boldsymbol{Z})$ given above in (9) with a bandwidth $h_m = m^{-\frac{1}{5}}$. After some simple algebra, the empirically infeasible test statistic can be computed as

$$CvM_{n0} \equiv \frac{1}{m} \sum_{j=1}^m \frac{1}{n^2} \boldsymbol{\psi}_j^\top \boldsymbol{P}_j \boldsymbol{P}_j^\top \boldsymbol{\psi}_j, \tag{20}$$

where for $j = 1, \ldots, m$,

$$\boldsymbol{\psi}_j \equiv (\psi_{\alpha_j}(\boldsymbol{W}_1, \boldsymbol{\beta}_0), \ldots, \psi_{\alpha_j}(\boldsymbol{W}_n, \boldsymbol{\beta}_0))^\top$$

---

[24]The asymptotic properties of the regression quantile estimator have been extensively investigated. In particular, it is known to satisfy the conditions of Assumption 2; see e.g., Gutenbrunner and Jurečková (1992, Theorem 1).

and

$$\boldsymbol{P}_j \equiv \boldsymbol{H}_j \boldsymbol{G},$$

where $\boldsymbol{H}_j \equiv \boldsymbol{I}_n - \boldsymbol{\delta}_j(\boldsymbol{\delta}_j^\top \boldsymbol{\delta}_j)^{-1}\boldsymbol{\delta}_j^\top$, $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix, $\boldsymbol{\delta}_j$ is the $n \times 3$ matrix whose ith row is denoted by

$$\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha_j))^\top = f\left(\boldsymbol{X}_i^\top \boldsymbol{\beta}_0(\alpha_j)\middle| \boldsymbol{X}_i\right)\boldsymbol{X}_i^\top$$

and $\boldsymbol{G}$ is the $n \times n$ matrix with elements $g_{ij} \equiv 1\left(\boldsymbol{X}_i \leq \boldsymbol{X}_j\right)$.

We denote by CvM_{n0}^* the bootstrapped analogue of CvM_{n0}, which is computed as

$$CvM_{n0}^* \equiv \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n^2}\boldsymbol{\psi}_j^{*\top}\boldsymbol{P}_j \boldsymbol{P}_j^\top \boldsymbol{\psi}_j^*,$$

where

$$\boldsymbol{\psi}_j^* \equiv (V_1\psi_{\alpha_j}(\boldsymbol{W}_1, \boldsymbol{\beta}_0(\alpha_j)), \ldots, V_n\psi_{\alpha_j}(\boldsymbol{W}_n, \boldsymbol{\beta}_0(\alpha_j)))^\top$$

and $\{V_i\}_{i=1}^{n}$ are iid Bernoulli random variates generated in accordance with the scheme given above in (16)–(17). These expressions indicate the computational simplicity of the proposed bootstrap procedure, as only the Bernoulli weights $\{V_i\}$ changes with each bootstrap replication.

The empirically feasible test statistic $\widehat{CvM}_n$ is computed in similar fashion to CvM_{n0} but with the parameter estimate $\hat{\boldsymbol{\beta}}_n(\cdot)$ replacing the true parameter vector $\boldsymbol{\beta}_0(\cdot)$ in the appropriate locations. In addition, the projection matrix $\boldsymbol{H}_j$ is replaced by $\hat{\boldsymbol{H}}_j \equiv \boldsymbol{I}_n - \hat{\boldsymbol{\delta}}_{nj}(\hat{\boldsymbol{\delta}}_{nj}^\top \hat{\boldsymbol{\delta}}_{nj})^{-1}\hat{\boldsymbol{\delta}}_{nj}^\top$, where $\hat{\boldsymbol{\delta}}_{nj}$ is an $n \times 3$ matrix with ith row denoted by

$$\hat{\boldsymbol{\delta}}_{nj}(\boldsymbol{X}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha_j))^\top = \hat{f}_{h_m}\left(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha_j)\middle| \boldsymbol{Z}_i\right)\boldsymbol{X}_i^\top.$$

The bootstrap approximation to $\widehat{CvM}_n$ is *mutatis mutandis* the same as its infeasible counterpart CvM_{n0}^*.

We compare our methods with subsampling-based tests (Chernozhukov and Fernández-Val, 2005, e.g.,) based on the non-projected empirical process given by S_n in (4) above.[25] The non-projected Cramér-von Mises test is given by

$$CvM_{nb0} \equiv \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n^2}\boldsymbol{\psi}_j^\top \tilde{\boldsymbol{X}} \tilde{\boldsymbol{X}}^\top \boldsymbol{\psi}_j, \tag{21}$$

---

[25]A comparison with a Cramér-von Mises test based on the Khmaladze (1981) transformation applied to S_n could also have been sensibly conducted in this context. It should also be noted that it is not possible to apply the multiplier bootstrap technique proposed in Section 4 to the simulation of critical values for tests based on continuous functionals of S_n. This is due to the presence of an additional $O_p(1)$ term in the asymptotic representation for S_n under the null that arises from

where $\tilde{\boldsymbol{X}}$ is the $n \times 3$ matrix with ith row given by $\tilde{\boldsymbol{X}}_i^\top = (\begin{array}{ccc} 1 & X_{1i} & X_{2i} \end{array})$.

Note that CvM_{nb0} differs only from the infeasible test statistic CvM_{n0} given above in (20) in the form of the matrix of the quadratic form in the summand. In particular, the true value of the parameter $\boldsymbol{\beta}_0$ is embedded in the test statistic as opposed to its estimate. We follow the suggestion of Sakov and Bickel (2000) applied to the m-out-of-n bootstrap and set the subsample size to $b = \left\lfloor \lambda n^{\frac{2}{5}} \right\rfloor$, where $\lambda > 0$. We experiment with a number of different settings for the leading constant λ and report the results for the setting of λ leading to the most accurately sized test, which for $n = 100$ yields an optimal subsample size of $b = 72.$[26] For simplicity, only subsamples consisting of blocks of consecutive observations in the original dataset were used in our simulations.

Table 1 displays rejection probabilities of the three tests considered under the data-generating process denoted by $DGP1$ in (18) above. The nominal sizes were set to 10%, 5% and 1%, while the values of the shift parameter c were taken to range from -.3 to .3, inclusive, in increments of .1. When $c = 0$, the results show that the size performance of our feasible test is accurate. The new bootstrap exhibits good size accuracy, uniformly across each of the nominal sizes considered, and for a sample size as small as $n = 100$. The distributional performance of the test based on $\widehat{CvM}_n$ is clearly superior to the test based on the conventional test based on subsampling. In particular, the subsampling-based test appears to perform especially poorly at controlling empirical sizes at a nominal level of 1%. In addition, the results of simulations not reported here indicate that the performance of the subsampling-based procedure is particularly sensitive to the choice of the subsample size b.

When $c \neq 0$, the results displayed in Table 1 indicate the power performance of the three tests considered. As one might expect, larger values of c lead to higher power for each of the tests. The tests based on multiplier-bootstrap approximations to CvM_{n0} and $\widehat{CvM}_n$ exhibit power much higher than the conventional test based on subsampling CvM_{nb0}. The infeasible and feasible tests appear to have similar power performance for $n = 100$, which indicates that little power is lost

---

the need to insert an estimate, rather than the true value, of the nuisance parameter $\boldsymbol{\beta}_0(\cdot)$ in the specification for S_n. This additional non-negligible term is not captured asymptotically when one attempts to apply the multiplier bootstrap technique described above to S_n. In this connection, see Theorem 5 below. This is a special case of the so-called "Durbin problem" (Durbin, 1973). Note by contrast the conclusion of Corollary 1—the orthogonal adjustment involved in the specification of $\hat{R}_n$ induces its weak convergence to a distribution-free mean-zero Gaussian process.

[26]Choosing the subsampling size optimally in this way is of course empirically infeasible as the true data-generating process is unknown.

when one replaces the true score with an estimate. This is striking given that only $m = 30$ observations are used to construct the nonparametric conditional density estimate given above in (9).

Table 1: Empirical rejection probabilities under $DGP1$

$n = 100$	CvM_{n0}			$\widehat{CvM}_n$			CvM_{nb0}		
c	10%	5%	1%	10%	5%	1%	10%	5%	1%
-.3	.997	.993	.950	.998	.992	.954	.671	.621	.531
-.2	.982	.949	.792	.979	.949	.778	.516	.456	.394
-.1	.739	.617	.349	.733	.620	.348	.301	.256	.204
0.0	.104	.051	.012	.104	.055	.012	.065	.044	.032
.1	.667	.519	.256	.670	.507	.247	.278	.245	.184
.2	.961	.913	.740	.956	.907	.719	.498	.443	.371
.3	.998	.987	.928	.995	.987	.928	.671	.628	.560

Table 2 displays empirical rejection probabilities for the three tests considered under DGP2 as given above in (19). The overall picture that emerges is qualitatively similar to that which emerges from Table 1. In particular, our proposed multiplier bootstrap-based tests exhibit more accurate empirical sizes and higher power than the conventional test based on subsampling.

Table 2: Empirical rejection probabilities under $DGP2$

$n = 100$	CvM_{n0}			$\widehat{CvM}_n$			CvM_{nb0}		
c	10%	5%	1%	10%	5%	1%	10%	5%	1%
-1.0	.640	.455	.146	.606	.409	.105	.223	.185	.132
-.5	.288	.164	.026	.263	.146	.021	.116	.087	.057
0.0	.090	.057	.008	.092	.055	.010	.074	.050	.032
.5	.517	.339	.092	.483	.295	.084	.181	.140	.101
1.0	.722	.524	.180	.691	.493	.174	.272	.229	.164

The simulations presented here have indicated that our proposals for multiplier bootstrap-based tests perform favourably when compared to more conventional specification tests based on subsampling. In particular, the results presented in Tables 1–2 indicate a substantial improvement in power associated with our proposals when compared with a specification test based on subsampling the statistic

24

given above in (21). Given the popularity of subsampling for inference involving the quantile regression process,[27] the results presented here suggest the possibility of less computationally demanding inference methodologies for the quantile regression process involving some form of multiplier bootstrap.

6   Empirical Example: Import Substitution in the Indian Pharmaceuticals Market

The desirability of patent enforcement on pharmaceuticals developed in high-income countries is often contentious when one considers its effect on consumer welfare in lower-income countries. This issue is bound up with the desirability of providing life-saving medicines to patients in developing countries at low cost—indeed, a typical argument made by governments of lower-income countries is that the enforcement of patents on essential medicines developed in rich countries will lead patients in poor countries to pay significantly more for these drugs than would otherwise be the case, leading in turn to adverse effects on the health and well-being of patients in lower-income countries. The other side of the debate typically involves the claim by multinational pharmaceutical firms that the enforcement of product patents is unlikely to have significant effects on prices due to the typical existence of lower-cost therapeutic substitutes for most patented drugs. A further claim is that the absence of effective patent protection in lower-income countries serves as a disincentive to basic research on diseases that have a disproportionate impact on patients in those countries. In other words, the enforcement of pharmaceutical product patents serves as a stimulus to product innovation.

In what follows we consider the structure of demand for a particular subsegment of the market for systemic anti-bacterial drugs (i.e., "antibiotics") in India. We make use of data originally analyzed by Chaudhuri et al. (2006) on sales involving antibiotics containing fluoroquinolone molecules. Chaudhuri et al. (2006) note that the Indian market for pharmaceutical products in the period between 1972 and 2005 provides an ideal setting for the study of the effects of global patent enforcement on consumer welfare in low-income countries. This is due on the one hand to the Indian government's non-recognition during this period of patents on phamaceutical products and on the other hand to the existence of a large domestic pharmaceutical industry with the capacity for producing and marketing

---

[27]e.g., Chernozhukov and Fernández-Val (2005).

drugs domestically that are under patent elsewhere.[28] The structure of demand in the Indian market is also similar to that of many other low-income countries because of the existence of a high proportion of uninsured households that are required to meet all expenses for drugs on an out-of-pocket basis. In addition, Chaudhuri et al. (2006) observe that the Indian pharmaceuticals market is typical of that of many lower-income countries due to the disproportionate importance of anti-infective drugs, which at 23 percent is the second-largest category in terms of overall market share.[29]

Chaudhuri et al. (2006) analyze a dataset that consists of monthly observations on sales of systemic antibiotic drugs in India over the period January 1999–December 2000.[30] The data are further disaggregated by geographical region, pharmaceutical product group and national origin (i.e., Indian or non-Indian), resulting in a total of 672 observations. Chaudhuri et al. (2006) focus on the particular market segment involving fluoroquinolone molecules, which denote a category of active pharmaceutical ingredients in treatments for a large number of different bacterial infections. [31] The fluoroquinolones segment is one of the largest in the Indian market for systemic antibiotics, accounting as it does for 20 percent of sales within this market. This segment is also characterized by the simultaneous availability in India during the sample period of antibiotic treatments that are protected by United States patents as well as of generic substitutes produced by domestic firms. The existence during the sample period of several close substitutes within the fluoroquinolones segment with different countries of origin enables the empirical evaluation of the claim that patent enforcement on foreign drugs raises prices in the domestic market. In particular, this claim is not credible if there exist significant substitution effects between patented and nonpatented drugs containing the same active pharmaceutical ingredient.

---

[28]Chaudhuri et al. (2006) note that pharmaceutical product patents were not recognized under Indian law between April 1972 and March 2005. They also note that the Indian pharmaceutical sector is now the world's largest producer, by volume, of generic formulations destined to be consumed by patients.

[29]Anti-infectives include both antibiotic and anti-viral drugs. This category is much more important in lower-income countries than in the overall world market for pharmaceuticals. In particular, anti-infectives account for only 9 percent of the worldwide market for pharmaceuticals.

[30]The actual dataset along with detailed information on each variable may be downloaded from `http://www.princeton.edu/~pennykg\TRIPS_Data&Programs.zip`.

[31]Chaudhuri et al. (2006, p. 1509).

6.1 The model

The empirical illustration presented here involves the estimation of a variant of the almost ideal demand system (AIDS) of Deaton and Muellbauer (1980b) involving two stages of expenditure allocation amongst categories of different pharmaceutical products.[32] The first stage involves a model of the optimal allocation of expenditures to various categories of systemic anti-bacterial drugs, including those containing fluoroquinolone molecules as active pharmaceutical ingredients. The second stage involves the optimal allocation of expenditures to the various product groups within the fluoroquinolones segment. In this connection, a fluoroquinolone "product group" is taken to refer to groups of pharmaceutical formulations produced by firms having the same national origin and containing the same active pharmaceutical ingredient within the fluoroquinolone category. The national origin of a pharmaceutical firm is taken to be one of two types, namely "domestic" if Indian or "foreign" if non-Indian, while the active pharmaceutical ingredient refers to a specific molecule within the fluoroquinolone family. This results in seven different fluoroquinolone product groups, to wit:

1. foreign ciprofloxacin

2. foreign norfloxacin

3. foreign ofloxacin

4. domestic ciprofloxacin

5. domestic norfloxacin

6. domestic ofloxacin

7. domestic sparfloxacin

The monthly data on prices and expenditures across each of these fluoroquinolone product groups are also disaggregated by geographical region, namely, "northern", "eastern", "western" or "southern".

In what follows, we focus specifically on an AIDS model for expenditure allocation amongst the first six of the seven product categories just listed. In particular, we examine the extent to which Indian consumers in the period January 1999–December 2000 engaged in behaviour consistent with substitution between foreign and domestic formulations containing the same fluoroquinolone molecule.

---

[32]See Deaton and Muellbauer (1980a, p. 131–132).

For a given fluoroquinolone product group, let

$$D^{01} \equiv \begin{cases} \text{the set of all other fluoroquinolone product groups with} \\ \qquad \text{different active pharmaceutical ingredients but} \\ \qquad \text{produced by firms of the same national origin} \end{cases}$$

$$D^{00} \equiv \begin{cases} \text{the set of all other fluoroquinolone product groups with} \\ \qquad \text{different active pharmaceutical ingredients and} \\ \qquad \text{produced by firms of different national origins} \end{cases}$$

For a given product group i within the fluoroquinolones category observed in geographical region r, let

$$\boldsymbol{p}_{ir} \equiv (p_{ir,11}, p_{ir,10}, \boldsymbol{p}_{ir,01}^\top, \boldsymbol{p}_{ir,00}^\top)^\top$$

be the relevant vector of prices, where

$$p_{ir,11} = \{ \text{ the own price}$$

$$p_{ir,10} = \begin{cases} \text{the price of the product group in the same region} \\ \text{having the same active pharmaceutical ingredient but} \\ \text{produced by firms of a different national origin.} \end{cases}$$

Also let

$$\boldsymbol{p}_{ir,01} \equiv (p_{jr,01} : j \in D^{01})^\top$$
$$\boldsymbol{p}_{ir,00} \equiv (p_{jr,00} : j \in D^{00})^\top$$

denote subvectors of prices for product groups in D^{01} and D^{00}, respectively.

The basic AIDS model we consider has the form

$$\begin{aligned} \log \omega_{ir\boldsymbol{p}_{ir}} = {} & \tau_{ir}(U_{ir}) + \gamma_{i,11}(U_{ir}) \log p_{ir,11} + \gamma_{i,10}(U_{ir}) \log p_{ir,10} \\ & + \gamma_{i,01}(U_{ir}) \sum_{j \in D^{01}} \log p_{jr,01} + \gamma_{i,00}(U_{ir}) \sum_{j \in D^{00}} \log p_{jr,00} \\ & + \beta_i(U_{ir}) \log \left(\frac{X_{Qr}}{P_{Qr}} \right), \end{aligned} \tag{22}$$

where $\omega_{ir\boldsymbol{p}_{ir}}$ is the expenditure share for product group i in region r when the vector of relevant prices is $\boldsymbol{p}_{ir}$, X_{Qr} is the overall expenditure in region r on fluoroquinolones, and P_{Qr} is the corresponding Stone price index. U_{ir} denotes an unobservable Uniform$(0,1)$ disturbance affecting expenditure shares of product

groups in each region, and $\tau_{ir}(U_{ir})$ is a constant term that captures product group-specific regional effects when the unobserved disturbance term is U_{ir}. Similarly, the parameters $\gamma_{i,11}(U_{ir})$, $\gamma_{i,10}(U_{ir})$, $\gamma_{i,01}(U_{ir})$, $\gamma_{i,00}(U_{ir})$ and $\beta_i(U_{ir})$ denote respectively the product group-specific own-price, cross-price and expenditure elasticities when the unobserved disturbance takes a value equal to U_{ir}. As such, a distinguishing characteristic of the AIDS specification (22) we consider for expenditure allocation within the fluoroquinolones category is that the price and expenditure elasticities are permitted to depend on the share of expenditures allocated to each product group. This allows for heterogeneity in the structure of demand for the various product groups according to their relative popularity in each region. In particular, patterns of substitutability between foreign and domestic formulations containing the same fluoroquinolone molecules are permitted to depend on the corresponding expenditure shares of the product groups in question.

By way of contrast, Chaudhuri et al. (2006, eq. (8)) consider a special case of the demand system given above in (22). In particular, they consider the location-shift variant of the general model given above in (22). Here the price and expenditure elasticities are constrained to be invariant across the distribution of expenditure shares.[33]

Time-series observations of the price vector $\boldsymbol{p}_{ir}$ for each product group i actually consist of price indices constructed in the manner described in Chaudhuri et al. (2006, p. 1492). The price index for each product group in the fluoroquinolones category involves weighting by the revenue share for each individual product, or "stock-keeping unit" (SKU) classified as belonging to the product group in question. These SKU revenue shares will generally depend on the overall product group revenue shares $\omega_{ir\boldsymbol{p}_{ir}}$ in (22), and as such will be correlated with the disturbance term U_{ir}. Here we follow Chaudhuri et al. (2006) in using instrumental variables to deal with the likely correlation of the price indices appearing in (22) with U_{ir}. In particular, the list of instrumental variables includes the number of SKUs in each product group, the prices of the five largest SKUs in each group according to revenue share and the natural logarithm of total expenditure on systemic anti-bacterial drugs for the time and region in question deflated using a Stone price index.[34]

---

[33]i.e., for each product group i, $\gamma_{i,11}(U_{ir})$, $\gamma_{i,10}(U_{ir})$, $\gamma_{i,00}(U_{ir})$ and $\beta_i(U_{ir})$ are assumed not to depend on U_{ir}.

[34]Further details on the instruments used appear in Chaudhuri et al. (2006, pp. 1492–1493).

6.2 Interpretation of the estimated cross-price elasticities

We focus on estimates of the parameter $\gamma_{i,10}(\cdot)$ appearing above in (22) for indices i corresponding to foreign ciprofloxacin, norfloxacin and ofloxacin on the one hand, and domestic ciprofloxacin, norfloxacin and ofloxacin on the other. In other words, we examine the substitutability of foreign and domestically produced formulations containing the same active pharmaceutical ingredient. Estimates of the cross-price elasticity parameters $\gamma_{i,10}$ for quantile indices in the range $[.05, .95]$ are plotted in Figure 1, along with the corresponding pointwise 95% confidence bands. These estimates along with their associated standard errors were implemented using the procedure of Blundell and Powell (2007) for censored regression quantiles with endogenous regressors, but with obvious simplifications made in the case of the present illustration to account for the lack of a censored dependent variable.[35]

The three graphs in the left-hand column of Figure 1 show the estimates of $\gamma_{i,10}$ for index i referring to the foreign product groups, while the three graphs in the right-hand column plot the corresponding parameter estimates for the domestic product groups. Figure 1 suggests a pattern of segmentation in the fluoroquinolones market between foreign and domestic formulations containing the same active pharmaceutical ingredient. In particular, expenditure shares for the three domestic product groups considered appear to exhibit uniformly little response across quantiles in $[.05, .95]$ to changes in the price levels of the foreign product groups containing the same active pharmaceutical ingredients. This characteristic appears to hold uniformly for all quantiles in the range considered, although the magnitude of the pointwise standard errors for domestic ciprofloxacin and norfloxacin indicate that any number of scenarios involving quantile dependence cannot be ruled out. The plots appearing in the right-hand column of Figure 1 suggest that if patent enforcement has any effect on the domestic prices of foreign anti-infective drugs, then that price effect has little impact on the market share of comparable domestic anti-infectives. This is a pattern that is most marked in the cases of domestic ciprofloxacin and ofloxacin. In summary, the right-hand column of Figure 1 suggests that one cannot rule out the possibility of patent enforcement on foreign antibiotic drugs leading to a higher overall price level in the

---

[35]A variant of the procedure of Blundell and Powell (2007) was adopted here instead of the better-known IVQR procedure of Chernozhukov and Hansen (2006) for reasons of practicality. In particular, the IVQR procedure is virtually infeasible with the number of endogenous covariates involved in the demand system described above. Further details and code are available from the authors on request.

Indian domestic market for antibiotics.

In contrast, the pattern that emerges down the left-hand column of Figure 1 suggests that domstic product groups exert at least a limited degree of price competition on foreign-produced formulations containing the same fluoroquinolone molecules. This is especially true for spending on foreign ciprofloxacin at quantiles below $\alpha = .8$ and for foreign norfloxacin at quantiles above $\alpha = .5$.

Finally, the apparently significant deviation from horizontality of most of the lines plotted in Figure 1 suggests that the location-shift variant of the basic AIDS model considered by Chaudhuri et al. (2006) may not be generally supported by the data available.

6.3   Tests of the AIDS model and of its location-shift specialization

In what follows, we report the results of two sets of diagnostic tests intended to evaluate the validity of the results reported above in the previous section. The first set of tests is more fundamental in nature and is intended to verify the functional form of the demand system given above in (22) for each of the product groups for which the cross-price elasticity estimates appear in Figure 1. In particular, if the assumed linearity in the relationship between log expenditure shares and the vector of log prices at all quantiles in $(0, 1)$ is not supported by the data, then the patterns of substitutability summarized in Figure 1 between foreign and domestic pharmaceutical alternatives containing the same active pharmaceutical ingredient cannot be reliably claimed as valid. In this connection, we verify the existence of a linear relationship between log expenditure shares and log prices using the methodology developed in the foregoing sections of this paper. In particular, we apply the same implementation of the empirically feasible Cramér-von Mises test described above in Section 5 to the data on each of the six product groups appearing in Figure 1. For each product group considered, the same variant of the Blundell and Powell (2007) technique described above was used to construct the estimates of the various demand-system parameters appearing in (22) when implementing the testing procedure. For each of the six product groups considered, the test was implemented in the same way and considered the hypothesis of linearity in parameters over quantiles in the range $[.05, .95]$. The results of the tests for each product group are indicated below in Table 3. These results indicate that the functional form of the demand system in (22) is supported overall by the data for each of the six product groups considered.

Table 3: Tests of linearity of structural quantile functions over the range $[.05, .95]$

Product group	p-value
Foreign ciprofloxacin	.5323
Foreign norfloxacin	.3682
Foreign ofloxacin	.5174
Domestic ciprofloxacin	.9950
Domestic norfloxacin	.9950
Domestic ofloxacin	.9950

The second set of tests is directed specifically at the validity of the model considered in the original paper of Chaudhuri et al. (2006). In particular, having accepted the validity of the general linear-in-parameters specification for the demand system given above in (22), we consider if the special case of the model characterized by constant price and expenditure elasticities over the range $(0, 1)$ is supported by the data at hand. In this connection, for each product group i, let

$$\boldsymbol{\theta}_i(U_{ir}) \equiv (\gamma_{i,11}(U_{ir}), \gamma_{i,10}(U_{ir}), \gamma_{i,00}(U_{ir}), \beta_i(U_{ir}))^\top,$$

where $\gamma_{i,11}$, $\gamma_{i,10}$, $\gamma_{i,00}$ and β_i are as given above in (22). We consider a test of the hypothesis

$$\boldsymbol{\theta}_i(\alpha) \equiv \boldsymbol{\theta}_i$$

for each i and each quantile α in the range $\mathcal{A} = [.05, .95]$. Several different approaches have already been proposed in the literature for testing this hypothesis, including the use of a Khmaladze martingale transformation (Bai, 2003; Koenker and Xiao, 2002) and subsampling (Chernozhukov and Fernández-Val, 2005). In what follows, we propose a variant of the approach suggested by Chernozhukov and Hansen (2006, Section 4).

For i corresponding to each of the six product groups considered, let $\hat{\boldsymbol{\theta}}_{ni}(\alpha)$ denote the corresponding estimator of $\boldsymbol{\theta}_i(\alpha)$ obtained using the same variant of the procedure of Blundell and Powell (2007) described above. Note that under the location-shift (i.e., constant coefficients) hypothesis, $\hat{\boldsymbol{\theta}}_{ni}(\alpha)$ should converge to the same value regardless of the quantile index α. Let $j = 1, \ldots, n_i$ denote the indices of the various observations for product group i; let $\boldsymbol{W}_{ji}$ denote the vector containing the log-expenditure share Y_{ji}, relevant prices and log expenditure on fluoroquinolones corresponding to the jth observation for product group i. Similarly, let $\boldsymbol{X}_{ji}$ denote the vector containing the relevant prices and log expenditure

on fluoroquinolones corresponding to the jth observation, and let $\boldsymbol{Z}_{ji}$ be the corresponding vector of instruments. Finally, let $\tilde{\boldsymbol{\theta}}_{ni}$ denote the two-stage least squares estimator of $\boldsymbol{\theta}_i$. In this connection, define the matrices

$$\boldsymbol{X}_i \equiv \begin{pmatrix} \boldsymbol{X}_{1i}^\top \\ \vdots \\ \boldsymbol{X}_{n_i i}^\top \end{pmatrix},$$

$$\boldsymbol{Z}_i \equiv \begin{pmatrix} \boldsymbol{Z}_{1i}^\top \\ \vdots \\ \boldsymbol{Z}_{n_i i}^\top \end{pmatrix}$$

and

$$\hat{\boldsymbol{X}}_i \equiv \left(\boldsymbol{X}_i^\top \boldsymbol{X}_i \right)^{-1} \boldsymbol{X}_i^\top \boldsymbol{Z}_i.$$

We exploit the asymptotic behaviour of a weighted empirical process corresponding to the Bahadur representation of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{ni}(\alpha) - \tilde{\boldsymbol{\theta}}_{ni} \right)$ that would obtain in the case where $\boldsymbol{\theta}_i(\alpha)$ is constant over $\alpha \in (0,1)$. In particular, consider the process given by

$$
\begin{aligned}
&\hat{Z}_{ni}(g, \alpha) \\
\equiv\ & \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big[\psi_\alpha(\boldsymbol{W}_{ji}, \hat{\boldsymbol{\theta}}_{ni}) \\
& \cdot \Big\{ g\left(\boldsymbol{Z}_{ji} \right) - \hat{\boldsymbol{D}}_n^\top(g, \hat{\boldsymbol{\theta}}_{ni}(\alpha)) \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_{ji}, \boldsymbol{X}_{ji}, \hat{\boldsymbol{\theta}}_{ni}(\alpha)) \Big\} \\
& - \left(Y_{ji} - \boldsymbol{X}_{ji}^\top \tilde{\boldsymbol{\theta}}_{ni} \right) \cdot g\left(\boldsymbol{Z}_{ji} \right) \Big\{ 1 - \hat{\boldsymbol{X}}_i^\top \left(\hat{\boldsymbol{X}}_i^\top \hat{\boldsymbol{X}}_i \right)^{-1} \hat{\boldsymbol{X}}_i \Big\} \Big], \quad (23)
\end{aligned}
$$

where $\psi_\alpha(\cdot, \cdot)$ is as given above in (2), $\hat{\boldsymbol{\delta}}_n$ and $\hat{\boldsymbol{\Delta}}_n$ are as given above in (8), and where in this case g denotes the indicator weighting function. As such, $\hat{Z}_{ni}$ is asymptotically free of the effects of estimating the nuisance parameter $\boldsymbol{\theta}_i$. Critical values for tests involving functionals of the form $\phi\left(\hat{Z}_{ni}(g, \cdot) \right)$ are simulated using the same multiplier bootstrap scheme described above in Section 4; in particular,

33

we make use of Monte Carlo approximations to the behaviour of the process

$$
\hat{Z}^*_{ni}(g, \alpha)
$$
$$
\equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big[\psi_\alpha(\boldsymbol{W}_{ji}, \hat{\boldsymbol{\theta}}_{ni})
$$
$$
\cdot \Big\{ g\left(\boldsymbol{Z}_{ji}\right) - \hat{\boldsymbol{D}}^\top_n (g, \hat{\boldsymbol{\theta}}_{ni}(\alpha)) \hat{\boldsymbol{\Delta}}^{-1}_n(\alpha) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_{ji}, \boldsymbol{X}_{ji}, \hat{\boldsymbol{\theta}}_{ni}(\alpha)) \Big\}
$$
$$
- \left(Y_{ji} - \boldsymbol{X}^\top_{ji} \tilde{\boldsymbol{\theta}}_{ni} \right) \cdot g\left(\boldsymbol{Z}_{ji}\right) \left\{ 1 - \hat{\boldsymbol{X}}^\top_i \left(\hat{\boldsymbol{X}}^\top_i \hat{\boldsymbol{X}}_i \right)^{-1} \hat{\boldsymbol{X}}_i \right\} \Big] V_j,
$$

where $\{V_j\}$ is a sequence of iid random variables with zero mean, unit variance, bounded support and also independent of each of the sequences $\{\boldsymbol{W}_{ji}\}$, $i = 1, \ldots, 6$. This allows for approximate critical values of tests based on functionals $\phi\left(\hat{Z}_{ni}(g, \cdot)\right)$ to be taken as appropriate empirical quantiles of simulated realizations of $\phi\left(\hat{Z}^*_{ni}(g, \cdot)\right)$.

Table 4: Tests for constant coefficients for quantiles over the range $[.05, .95]$

Product group	p-value
Foreign ciprofloxacin	.5075
Foreign norfloxacin	.0398
Foreign ofloxacin	.1965
Domestic ciprofloxacin	.0100
Domestic norfloxacin	.0050
Domestic ofloxacin	.0149

Table 4 reports the results of tests of hypotheses of constant price and expenditure elasticities over the range of quantiles $[.05, .95]$ for each of the six fluoroquinolone product groups considered in Figure 1. Each of the rows in the table corresponds to implementations of Cramér-von Mises tests based on the process given above in (23) for a given product group implemented with simulated sequences of bootstrap multipliers $\{V_j\}$ generated according to the scheme given above in (16)–(17). Each of these product group-specific tests was implemented in the same way. In particular, for each row of Table 4, the integrals characterizing the Cramér-von Mises statistics were approximated by averages of quadratic forms taken over grids of thirty evenly spaced quantiles in the range $[.05, .95]$.

This was done in the same manner described above for the simulation experiments reported in Section 5. The conditional density estimates embedded in the expressions for $\hat{\boldsymbol{\delta}}_n$ and $\hat{\boldsymbol{\Delta}}_n$ in (23) were also constructed in the same manner described above for the simulation experiments reported in Section 5.

The tests reported in Table 4 indicate that for most of the product groups considered, the location-shift specialization of the basic demand system given above in (22) is not supported by the data at hand. In particular, the demand systems corresponding to the three domestic product groups considered are not well described by linear-in-parameters models with constant coefficients. This conclusion underscores the extent to which a consideration of the more flexible demand system specification given above in (22) can provide a more complete picture of the nature of price competition in a system of related goods than the more common specification with constant coefficients adopted by Chaudhuri et al. (2006).

7   Conclusion

This paper has proposed a class of tests for the hypothesis of linearity in parameters of measurable functions that are identified by conditional quantile restrictions involving instrumental variables. We have argued that these tests provide a potentially useful diagnostic tool for empirical researchers interested in the effect of possibly endogenous conditioning variables on the distribution of some outcome variable of interest. We exploit various results from empirical process theory to derive the asymptotic behaviour of our proposed test procedures. In a manner analogous to approaches taken by Neyman (1959) and Bickel et al. (2006) to problems that are qualitatively similar to ours, we have also shown how the use of an orthogonal projection on the tangent space of nuisance parameters at each quantile improves power performance while at the same time serving to facilitate the simulation of critical values via the application of a simple multiplier bootstrap procedure. Simulation evidence and an empirical example involving the structure of demand for anti-bacterial drugs in India illustrate the feasibility of our approach in datasets of moderate size.

References

Amemiya, T. (1982) 'Two stage least absolute deviations estimators.' *Econometrica* 50, 689–711

Andrews, D. W. K. (1997) 'A conditional Kolmogorov test.' *Econometrica* 65, 1097–1128

Bai, J. (2003) 'Testing parametric conditional distributions of dynamic models.' *Review of Economics and Statistics* 85, 531–549

Bickel, P. J., Y. Ritov, and T. M. Stoker (2006) 'Tailor-made tests for goodness of fit to semiparametric hypotheses.' *Annals of Statistics* 34, 721–741

Bierens, H. J., and D. K. Ginther (2001) 'Integrated conditional moment testing of quantile regression models.' *Empirical Economics* 26, 307–324

Bierens, H. J., and W. Ploberger (1997) 'Asymptotic theory of integrated conditional moment tests.' *Econometrica* 65, 1129–1151

Bilias, Y., S. Chen, and Z. Ying (2000) 'Simple resampling methods for censored regression quantiles.' *Journal of Econometrics* 99, 373–386

Blundell, R. W., and J. L. Powell (2007) 'Censored regression quantiles with endogenous regressors.' *Journal of Econometrics* 141, 65–83

Chaudhuri, S., P. K. Goldberg, and P. Jia (2006) 'Estimating the effects of global patent protection in pharmaceuticals: A case study of quinolones in india.' *American Economic Review* 96(5), 1477–1513

Chen, L., and S. Portnoy (1996) 'Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models.' *Communications in Statistics, Theory and Methods* 25, 1005–1032

Chen, X., O. Linton, and I. van Keilegom (2003) 'Estimation of semiparametric models when the criterion function is not smooth.' *Econometrica* 71, 1591–1608

Chernozhukov, V., and C. Hansen (2004) 'The effects of 401(k) participation on the wealth distribution: An instrumental quantile regression analysis.' *Review of Economics and Statistics* 86, 735–751

___ (2005) 'An IV model of quantile treatment effects.' *Econometrica* 73, 245–262

___ (2006) 'Instrumental quantile regression inference for structural and treatment effect models.' *Journal of Econometrics* 132, 491–525

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

Chernozhukov, V., I. Fernández-Val, and B. Melly (2009) 'Inference on counterfactual distributions.' cemmap Working Paper CWP09/09, Institute for Fiscal Studies

Deaton, A., and J. Muellbauer (1980a) *Economics and Consumer Behavior* (Cambridge, U.K.: Cambridge University Press)

Deaton, A. S., and J. Muellbauer (1980b) 'An almost ideal demand system.' *American Economic Review* 70(3), 312–326

Dudley, R. M. (1999) *Uniform Central Limit Theorems* (Cambridge, U.K.: Cambridge University Press)

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

Einmahl, U., and D. M. Mason (2005) 'Uniform in bandwidth consistency of kernel-type function estimators.' *Annals of Statistics* 33, 1380–1403

Escanciano, J. C. (2007) 'Weak convergence of non-stationary multivariate marked processes with applications to martingale testing.' *Journal of Multivariate Analysis* 98, 1321–1336

— (2009) 'On the lack of power of omnibus specifiaction tests.' *Econometric Theory* 25, 162–194

Escanciano, J. C., and C. Velasco (2006) 'Specification tests of parametric dynamic conditional quantiles.' Department of Economics, Indiana University

Forbes, S. J. (2008) 'The effect of air traffic delays on airline prices.' *International Journal of Industrial Organization* 26, 1218–1232

Giné, E., and J. Zinn (1990) 'Bootstrapping general empirical measures.' *Annals of Probability* 18, 851–869

— (1991) 'Gaussian characterization of uniform Donsker classes of functions.' *Annals of Probability* 19, 758–782

Gutenbrunner, C., and J. Jurečková (1992) 'Regression rank scores and regression quantiles.' *Annals of Statistics* 20, 305–330

Hahn, J. (1995) 'Bootstrapping quantile regression models.' *Econometric Theory* 11, 105–121

He, X., and F. Hu (2002) 'Markov chain marginal bootstrap.' *Journal of the American Statistical Association* 97, 783–795

Honoré, B. E., and L. Hu (2004) 'On the performance of some robust instrumental variables estimators.' *Journal of Business and Economic Statistics* 22, 30–39

Horowitz, J. L. (1998) 'Bootstrap methods for median regression models.' *Econometrica* 66, 1327–1351

Horowitz, J. L., and S. Lee (2009) 'Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative.' *Journal of Econometrics* 152, 141–152

Horowitz, J. L., and V. G. Spokoiny (2002) 'An adaptive, rate-optimal test of linearity for median regression models.' *Journal of the American Statistical Association* 97, 822–835

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

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

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

Lee, S. (2007) 'Endogeneity in quantile regression models: A control function approach.' *Journal of Econometrics* 141, 1131–1158

Ma, L., and R. Koenker (2006) 'Quantile regression methods for recursive structural equation models.' *Journal of Econometrics* 134, 471–506

Machado, J. A. F., and J. Mata (2005) 'Counterfactual decomposition of changes in wage distributions using quantile regression.' *Journal of Applied Econometrics* 20, 445–465

Neyman, J. (1959) 'Optimal asymptotic tests of composite statistical hypotheses.' In *Probability and Statistics: The Harald Cramér Volume,* ed. U. Grenander (Stockholm: Almqvist and Wiksell) pp. 213–234

Parzen, E. (1962) 'On estimation of a probability density function and mode.' *Annals of Statistics* 33, 1065–1076

Powell, J. L. (1983) 'The asymptotic normality of two-stage least absolute deviations estimators.' *Econometrica* 51, 1569–1576

Rosenblatt, M. (1956) 'Remarks on some nonparametric estimates of a density function.' *Annals of Mathematical Statistics* 27, 832–837

Sakata, S. (2007) 'Instrumental variable estimation based on conditional median restriction.' *Journal of Econometrics* 141, 350–382

Sakov, A., and P. J. Bickel (2000) 'An Edgeworth expansion for the m out of n bootstrapped median.' *Statistics and Probability Letters* 49, 217–223

Stinchcombe, M., and H. White (1998) 'Consistent specification testing with nuisance parameters present only under the alternative.' *Econometric Theory* 14, 295–325

Strasser, H. (1985) *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory* (Berlin: Walter de Gruyter)

Stute, W. (1997) 'Nonparametric model checks for regression.' *Annals of Statistics* 25, 613–641

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

Whang, Y.-J. (2006a) 'Consistent specification testing for quantile regression models.' In *Econometric Theory and Practice: Frontiers of Analysis and Applied Research,* ed. D. Corbae, S. N. Durlauf, and B. E. Hansen (New York: Cambridge University Press) pp. 288–308

— (2006b) 'Smoothed empirical likelihood methods for quantile regression models.' *Econometric Theory* 22, 173–205

Wu, C. F. J. (1986) 'Jackknife, bootstrap and other resampling methods in regression analysis.' *Annals of Statistics* 14, 1261–1350. with discussion

Zheng, J. X. (1998) 'A consistent nonparametric test of parametric regression models under conditional quantile restrictions.' *Econometric Theory* 14, 123–138

A   Appendix

A.1   Preliminary results

We begin with an important result of Chen et al. (2003) that allows for the bounding of entropy numbers and the verification of stochastic equicontinuity for processes indexed by both Euclidean and function-valued parameters. In this connection, define a generic function class

$$\mathcal{H} = \{\boldsymbol{t} \to m(\boldsymbol{t}, \boldsymbol{\theta}, g) : \boldsymbol{\theta} \in \Theta, g \in \mathcal{G}\},$$

where Θ and $\mathcal{G}$ are generic Banach spaces with associated norms $\|\cdot\|_\Theta$ and $\|\cdot\|_\mathcal{G}$, respectively. Recall that the *covering number* $N(\epsilon, \Theta, \|\cdot\|_\Theta)$ of Θ is the minimal number N for which there exist ϵ-neighbourhoods $\{\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|_\Theta \leq \epsilon\}, \|\boldsymbol{\theta}_j\|_\Theta < \infty, j = 1, \dots, N\}$ covering Θ. The *covering number with bracketing* $N_{[]}(\epsilon, \Theta, \|\cdot\|_\Theta)$ is the minimal number N for which there exist ϵ-brackets $\{(\boldsymbol{l}_j, \boldsymbol{u}_j) : \|\boldsymbol{l}_j - \boldsymbol{u}_j\|_\Theta \leq \epsilon, \|\boldsymbol{l}_j\|_\Theta, \|\boldsymbol{u}_j\|_\Theta < \infty, j = 1, \dots, N\}$ covering Θ. In particular, the idea here is that for each $\boldsymbol{\theta} \in \Theta$, there is a $j = j_{\boldsymbol{\theta}} \in \{1, \dots, N\}$ such that $\boldsymbol{l}_j \leq \boldsymbol{\theta} \leq \boldsymbol{u}_j$. Other definitions of concepts from empirical process theory may be found in e.g., van der Vaart and Wellner (1996).

Lemma 1 (Chen, Linton & van Keilegom (2003, Theorem 3)).   *1. Assume that*

$$|m(\boldsymbol{t}, \boldsymbol{\theta}_1, g_1) - m(\boldsymbol{t}, \boldsymbol{\theta}_2, g_2)| \leq b(\boldsymbol{t}) \left\{ \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_\Theta^{s_1} + \|g_1 - g_2\|_\mathcal{G}^{s_2} \right\} \tag{24}$$

for some constants $s_1, s_2 \in (0, 1]$ and for some measurable function b with $\|b\|_{r,P} < \infty$, where $r \geq 2$. Then for any $\epsilon > 0$, thhe covering number with bracketing of the class $\mathcal{H}$ satisfies

$$N_{[\cdot]}(\epsilon, \mathcal{H}, \|\cdot\|_{r,P}) \leq N\left(\left[\frac{\epsilon}{4\|b\|_{r,P}}\right]^{\frac{1}{s_1}}, \Theta, \|\cdot\|_\Theta\right) \times N\left(\left[\frac{\epsilon}{4\|b\|_{r,P}}\right]^{\frac{1}{s_2}}, \mathcal{G}, \|\cdot\|_\mathcal{G}\right).$$

2. Assume that

$$E\left[\sup_{\boldsymbol{\theta}_2: \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_\Theta < \delta} \sup_{g_2: \|g_1 - g_2\|_\mathcal{G} < \delta} |m(\boldsymbol{t}, \boldsymbol{\theta}_1, g_1) - m(\boldsymbol{t}, \boldsymbol{\theta}_2, g_2)|^r\right] \leq K\delta^{rs_1} \tag{25}$$

for some constant $s_1 \in (0, 1]$ and for some $r \geq 2$. Then for any $\epsilon > 0$,

$$N_{[\cdot]}(\epsilon, \mathcal{H}, \|\cdot\|_{r,P}) \leq N\left(\left[\frac{\epsilon}{2K}\right]^{\frac{1}{s_1}}, \Theta, \|\cdot\|_{\Theta}\right) \times N\left(\left[\frac{\epsilon}{2K}\right]^{\frac{1}{s_1}}, \mathcal{G}, \|\cdot\|_{\mathcal{G}}\right).$$

Suppose that conditions (24) and (25) both hold. Then if Θ is a compact subset of $\mathbb{R}^k$ for some k, and if

$$\int_0^\infty \sqrt{\log N\left(\epsilon^{\frac{1}{s_j}}, \mathcal{G}, \|\cdot\|_{\mathcal{G}}\right)} d\epsilon < \infty$$

for $j = 1, \ldots, l$, the empirical process $\left\{M_n(\boldsymbol{\theta}, g) \equiv \frac{1}{n}\sum_{i=1}^n m\left(\boldsymbol{T}_i, \boldsymbol{\theta}, g\right) : \boldsymbol{\theta} \in \Theta, g \in \mathcal{G}\right\}$ is asymptotically stochastically equicontinuous, i.e., for any sequence of positive constants $\delta_n = o(1)$,

$$\sup_{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_\Theta \leq \delta_n, \|g_1 - g_2\|_{\mathcal{G}} \leq \delta_n} \|M_n(\boldsymbol{\theta}_1, g_1) - M_n(\boldsymbol{\theta}_2, g_2) - M(\boldsymbol{\theta}_1, g_1)\| = o_p(n^{-\frac{1}{2}}),$$

where $M(\boldsymbol{\theta}_1, g_1) \equiv E\left[m(\boldsymbol{T}_i, \boldsymbol{\theta}_1, g_1)\right]$.

We now state a weak convergence theorem that is useful in dealing with estimation effects in test functionals involving the non-smooth summands $\psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n)$. This result is of independent interest. Define the function-weighted empirical process

$$V_n(\boldsymbol{\beta}, \alpha, g) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}) - E\left[\psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}) | \boldsymbol{Z}_i\right]\right) g(\boldsymbol{Z}_i),$$

which is indexed by $\gamma \equiv (\boldsymbol{\beta}, \alpha, g) \in \mathcal{F} \times \mathcal{A} \times \mathcal{G}$, where $\mathcal{F}$ is the class of measurable $\mathbb{R}^d$-valued functions of $(0, 1)$ to which $\boldsymbol{\beta}$ is assumed to belong, $\mathcal{A}$ is either $[0, 1]$ or a compact subset of $(0, 1)$ and $\mathcal{G}$ is taken to be a class of measurable functions with a measurable envelope $G(\boldsymbol{Z})$.

Theorem 4. *Under the conditions of Assumptions 1–2, the process $V_n(\gamma)$ is stochastically equicontinuous with respect to $\|\cdot\|_{2,P}$.*

Proof. We shall apply van der Vaart and Wellner (1996, Theorem 2.11.23). For $\gamma \equiv (\boldsymbol{\beta}, \alpha, g)$ and $\boldsymbol{w} \equiv (\boldsymbol{x}^\top, y)^\top$, define the function class $\mathcal{H} \equiv \{\boldsymbol{w} \to h(\boldsymbol{w}, \gamma)\}$, where

$$h(\boldsymbol{w}, \gamma) \equiv (\psi_\alpha(\boldsymbol{w}, \boldsymbol{\beta}) - E\left[\psi_\alpha(\boldsymbol{W}, \boldsymbol{\beta}) | \boldsymbol{Z} = \boldsymbol{z}\right]) g(\boldsymbol{z}).$$

Fix $\gamma_1 \equiv (\boldsymbol{\beta}_1, \alpha_1, g_1) \in \mathcal{F} \times \mathcal{A} \times \mathcal{G}$. Let δ be a constant in $(0, 1)$ and define $\boldsymbol{\iota}$ to be the vector $(1, 1, ..., 1)^\top \in \mathbb{R}^d$. By the triangle inequality and recalling Assumption 2, we have

$$E\left[\sup_{\boldsymbol{\beta}: \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}\|_\mathcal{F} < \delta} \sup_{\alpha: |\alpha_1 - \alpha| < \delta} \sup_{g_2: \|g_1 - g_2\|_\mathcal{G} < \delta} |h(\boldsymbol{W}, \gamma_1) - h(\boldsymbol{W}, \gamma_2)|^2\right]$$

$$\leq CE\left[\sup_{\alpha: |\alpha_1 - \alpha| < \delta} \sup_{\boldsymbol{\beta}: \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}\|_\mathcal{F} < \delta} |E\left[\psi_{\alpha_1}(\boldsymbol{W}, \boldsymbol{\beta}_1) | \boldsymbol{Z}\right] - E\left[\psi_\alpha(\boldsymbol{W}, \boldsymbol{\beta}) | \boldsymbol{Z}\right]|^2\right] + C\delta^{\frac{1}{2}}$$

$$\leq CE\left[\sup_{\alpha_1 \in \mathcal{A}} \left(F\left(\boldsymbol{X}^\top\boldsymbol{\beta}_1(\alpha_1) + \boldsymbol{X}^\top\boldsymbol{\iota}\delta \big| \boldsymbol{Z}\right) - F\left(\boldsymbol{X}^\top\boldsymbol{\beta}_1(\alpha_1) - \boldsymbol{X}^\top\boldsymbol{\iota}\delta\big| \boldsymbol{Z}\right)\right)^2\right] + C\delta^{\frac{1}{2}}$$

$$\leq C\delta.$$

Apply Lemma 1. The desired conclusion follows from van der Vaart and Wellner (1996, Theorem 2.11.23). □

Next, consider the asymptotic behaviour under H_0 of the process

$$S_n(g, \alpha) = n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n) g(\boldsymbol{Z}_i)$$

as defined earlier in (4) above. We have the following result, which follows immediately from the conclusion of Theorem 4.

Theorem 5. *The following representation holds uniformly for* $(\boldsymbol{\beta}, \alpha, g) \in \mathcal{F} \times \mathcal{A} \times \mathcal{G}$ *under the conditions of Assumptions 1–3:*

$$
\begin{aligned}
S_n(g, \alpha) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(F\left(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha) \Big| \boldsymbol{Z}_i \right) - F\left(\boldsymbol{X}_i^\top \boldsymbol{\beta}_0(\alpha) \Big| \boldsymbol{Z}_i \right) \right) g(\boldsymbol{Z}_i) + o_p(1).
\end{aligned}
$$

Proof. Apply Theorem 4 and conclude that

$$\sup_{(\alpha, g) \in \mathcal{A} \times \mathcal{G}} \left| V_n(\hat{\boldsymbol{\beta}}_n, \alpha, g) - V_n(\boldsymbol{\beta}_0, \alpha, g) \right| = o_p(1),$$

which is equivalent to

$$
\begin{aligned}
\sup_{(\alpha, g) \in \mathcal{A} \times \mathcal{G}} &\left| S_n(g, \alpha) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i) \right. \\
&\left. + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(E\left[\psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) | \boldsymbol{Z}_i \right] - E\left[\psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n) \Big| \boldsymbol{Z}_i \right] \right) g(\boldsymbol{Z}_i) \right| = o_p(1).
\end{aligned}
$$

This in turn may be rewritten as the desired conclusion. □

In what follows we make use of a general result for uniform convergence of a kernel estimator over classes of functions and bandwidths restricted to lie in suitable bounded intervals. In this connection, a preliminary result is needed. Let $\Psi(\cdot)$ be a bounded non-decreasing and right-continuous function. Consider a class of real-valued measurable functions $\mathcal{T}$ on $\mathbb{R}^d \times [0, 1]$, and define the function class

$$\mathcal{H} \equiv \left\{ \Psi\left(\frac{u - \tau(\cdot)}{h} \right) : h \in (0, 1], u \in \mathbb{R}, \tau \in \mathcal{T} \right\}.$$

We have the following.

Lemma 2. *Suppose that the linear span of* $\mathcal{T}$ *is a finite-dimensional set of functions. Then* $\mathcal{H}$ *is a VC-subgraph class.*

41

Proof. For any bounded non-decreasing right-continuous function Ψ, define for $t \in \mathbb{R}$ the left-continuous inverse of Ψ as

$$\Psi^{-1}(t) \equiv \inf \left\{ v : \; \Psi(v) \geq t \right\},$$

where the convention $\inf \emptyset \equiv \infty$ is observed. Since $\Psi(\cdot)$ is non-decreasing and right-continuous we have for any $t \in \mathbb{R}$

$$\Psi(v) < t \Leftrightarrow v < \Psi^{-1}(t). \tag{26}$$

For fixed constants $h \in (0, 1]$ and $u \in \mathbb{R}$, define the class of subgraphs

$$\mathcal{S}_{u,h} \equiv \left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; \Psi \left(\frac{u - \tau(\boldsymbol{x}, \alpha)}{h} \right) < t : \; \Psi \in \mathcal{H} \right\}.$$

We need to show that each $\mathcal{S}_{u,h}$ is a VC-class.[36] By virtue of (26), we have

$$
\begin{aligned}
\mathcal{S}_{u,h} &= \left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; \frac{u - \tau(\boldsymbol{x}, \alpha)}{h} < \Psi^{-1}(t) \right\} \\
&= \left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; \tau(\boldsymbol{x}, \alpha) - u + h\Psi^{-1}(t) > 0 \right\}.
\end{aligned} \tag{27}
$$

Let $t_+ \equiv \inf \left\{ t : \; \Psi^{-1}(t) = \infty \right\}$, and $t_- \equiv \sup \left\{ t : \; \Psi^{-1}(t) = -\infty \right\}$. We note that $\Psi^{-1}(t)$ is finite iff $t \in (t_-, t_+)$. Let ϕ denote the restriction of Ψ^{-1} to (t_-, t_+). Let $\mathcal{F}$ be the linear span of $\mathcal{T} \cup \{1, \phi(\cdot)\}$. Each function $f \in \mathcal{F}$ is defined on $\mathbb{R}^d \times [0, 1] \times (t_-, t_+)$ and has the form

$$f(\boldsymbol{x}, \alpha, t) = c_1 \tau(\boldsymbol{x}, \alpha) + c_2 + c_3 \phi(t)$$

for $(\boldsymbol{x}^\top, \alpha, t)^\top \in \mathbb{R}^d \times [0, 1] \times (t_-, t_+)$, $t \in \mathcal{T}$ and $c_1, c_2, c_3 \in \mathbb{R}$.

Since the linear span of $\mathcal{T}$ is finite dimensional, it follows that $\mathcal{F}$ is also finite dimensional. It follows from van der Vaart and Wellner (1996, Lemma 2.6.15) that $\mathcal{F}$ is a VC-subgraph class.

Note that for every $\tau \in \mathcal{T}$, $u \in \mathbb{R}$ and $h > 0$,

$$
\begin{aligned}
&\left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; \tau(\boldsymbol{x}, \alpha) + u + h\Psi^{-1}(t) > 0 \right\} \\
&= \left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; \tau(\boldsymbol{x}, \alpha) + u + h\phi(t) > 0, \, t \in (t_-, t_+) \right\} \cup \mathbb{R}^d \times [t_+, \infty).
\end{aligned}
$$

It follows from (27) that

$$\mathcal{S}_{u,h} \subset \left\{ \left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; f(\boldsymbol{x}, \alpha, t) > 0 \right\} \cup \mathbb{R}^d \times [t_+, \infty) : \; f \in \mathcal{F} \right\} \equiv \mathcal{D}. \tag{28}$$

Note that $\mathbb{R}^d \times [t_+, \infty)$ is trivially a VC-class. In addition, we have by van der Vaart and Wellner (1996, Lemma 2.6.18(iii)) that the class

$$\left\{ \left\{ (\boldsymbol{x}^\top, \alpha, t)^\top : \; f(\boldsymbol{x}, \alpha, t) > 0 \right\} : \; f \in \mathcal{F} \right\}$$

is also VC. Then by van der Vaart and Wellner (1996, Lemma 2.6.17(iii)) we have that $\mathcal{D}$ is VC, which by the inclusion in (28) implies that $\mathcal{H}$ is a VC class. $\qquad \square$

---

[36]Cf. e.g., van der Vaart and Wellner (1996, p. 141).

Now consider the class of functions

$$\mathcal{G} \equiv \left\{ K\left(\frac{u - \tau(\cdot)}{h}\right) : h \in (0, 1], u \in \mathbb{R}, \tau \in \mathcal{T} \right\},$$

where $K(\cdot)$ satisfies the conditions of Assumption 4 and τ the corresponding condition of Lemma 2. Under the conditions of Assumption 4, we have the following slight modification of the uniform-in-bandwidth consistency result for a kernel density estimator given in Einmahl and Mason (2005, Theorem 1):

Lemma 3. *Suppose $K(\cdot)$ satisfies the conditions of Assumption 4, while $\tau(\cdot)$ is as in the statement of Lemma 2 above. Then for any sequence $\{(a_m, b_m) : m = 1, 2, \ldots\}$ satisfying*

$$0 < a_m < b_m \leq 1$$

with $b_m \to 0$, $\frac{m a_m}{\log m} \to \infty$ as $m \to \infty$, we have

$$\sup_{a_m \leq h \leq b_m} \sup_{u \in \mathbb{R}} \sup_{\tau \in \mathcal{T}} \left| \frac{1}{mh} \sum_{j=1}^{m} K\left(\frac{u - \tau(\boldsymbol{X}, U_j)}{h}\right) - \frac{1}{h} E\left[K\left(\frac{u - \tau(\boldsymbol{X}, U)}{h}\right)\right] \right|$$

$$= O\left(\sqrt{\frac{\log\left(\frac{1}{a_m}\right) \vee \log \log m}{m a_m}}\right)$$

almost surely as $m \to \infty$.

Proof. We note from Lemma 2 that

$$\mathcal{H}_i \equiv \left\{ \Psi_i\left(\frac{u - \tau(\cdot)}{h}\right) : h \in (0, 1], u \in \mathbb{R}, \tau \in \mathcal{T} \right\}$$

is a VC-subgraph class for $i = 1, 2$. By van der Vaart and Wellner (1996, Theorem 2.6.7), it follows that for each $i \in \{1, 2\}$, there exists $C_i > 0$ and $\nu > 0$ such that

$$N(\epsilon, \mathcal{H}_i) \leq C_i \epsilon^{\nu_i}$$

for $\epsilon \in (0, 1)$ and each $i \in \{1, 2\}$. From this it follows that for some $C > 0$ and $\nu > 0$,

$$N(\epsilon, \mathcal{G}) \leq C \epsilon^{\nu}.$$

By virtue of the right-continuity of K assumed in Assumption 4, we have that $\mathcal{G}$ is pointwise measurable, i.e., there is a countable subclass $\mathcal{G}_0$ of $\mathcal{G}$ such that for every $g \in \mathcal{G}$, there exists a sequence of functions $\{g_m\} \subset \mathcal{G}_0$ such that for each $u \in \mathbb{R}$,

$$g_m(u) \to g(u).$$

The desired conclusion follows from the conditions of Assumption 4 and the proof of Einmahl and Mason (2005, Theorem 1). $\qquad\square$

Lemma 3 is exploited to yield the following uniform consistency result for the conditional density estimator given above in (9).

Lemma 4. *Let $\mathcal{A}$ denote a compact subset of $[0, 1]$. Under the conditions of Assumptions 1–4 and Lemma 3 we have*

$$\sup_{a_m \leq h \leq b_m} \sup_{\boldsymbol{z} \in \mathbb{R}^k} \sup_{\boldsymbol{x} \in \mathbb{R}^d} \sup_{\alpha \in \mathcal{A}} \left| \hat{f}_h \left(\boldsymbol{x}^\top \hat{\boldsymbol{\beta}}_n(\alpha) \middle| \boldsymbol{z} \right) - f \left(\boldsymbol{x}^\top \boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{z} \right) \right|$$

$$= O_p \left(\frac{1}{a_m^2 \sqrt{n}} + \sqrt{\frac{\log \left(\frac{1}{a_m} \right) \vee \log \log m}{m a_m}} + b_m^2 \right)$$

as $m, n \to \infty$.

Proof. Invoke the conclusion of Lemma 3 and the Lipschitz condition on K assumed in Assumption 4 to deduce the following bound holding uniformly for $\boldsymbol{\beta} \in \mathcal{F}$ as $m \to \infty$:

$$\left| \frac{1}{mh} \sum_{j=1}^m \left(K \left(\frac{u - \boldsymbol{X}^\top \boldsymbol{\beta}(U_j)}{h} \right) - K \left(\frac{u - \boldsymbol{X}^\top \boldsymbol{\beta}_0(U_j)}{h} \right) \right) \right| \leq \frac{M_0 \|\boldsymbol{X}\| \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|}{h^2} \quad (29)$$

almost surely for some $M_0 > 0$.

Combining the conclusion of Lemma 3, (29) and the condition of Assumption 3 we have

$$\sup_{a_m \leq h \leq b_m} \sup_{u \in \mathbb{R}} \left| \frac{1}{mh} \sum_{j=1}^m K \left(\frac{u - \boldsymbol{X}^\top \hat{\boldsymbol{\beta}}_n(U_j)}{h} \right) - \frac{1}{h} E \left[K \left(\frac{u - \boldsymbol{X}^\top \boldsymbol{\beta}_0(U)}{h} \right) \right] \right|$$

$$\leq \sup_{a_m \leq h \leq b_m} \sup_{u \in \mathbb{R}} \left| \frac{1}{mh} \sum_{j=1}^m \left(K \left(\frac{u - \boldsymbol{X}^\top \hat{\boldsymbol{\beta}}_n(U_j)}{h} \right) - K \left(\frac{u - \boldsymbol{X}^\top \boldsymbol{\beta}_0(U_j)}{h} \right) \right) \right|$$

$$+ \sup_{a_m \leq h \leq b_m} \sup_{u \in \mathbb{R}} \sup_{\tau \in \mathcal{T}} \left| \frac{1}{mh} \sum_{j=1}^m K \left(\frac{u - \tau(\boldsymbol{X}, U_j)}{h} \right) - \frac{1}{h} E \left[K \left(\frac{u - \tau(\boldsymbol{X}, U)}{h} \right) \right] \right|$$

$$\leq O_p \left(\frac{1}{a_m^2 \sqrt{n}} \right) + O_p \left(\sqrt{\frac{\log \left(\frac{1}{a_m} \right) \vee \log \log m}{m a_m}} \right). \quad (30)$$

Now fix $\boldsymbol{z} \in \mathbb{R}^k$ and $\boldsymbol{x} \in \mathbb{R}^d$. Let $W_{\boldsymbol{x}} \equiv \boldsymbol{x}^\top \boldsymbol{\beta}_0(U)$. We recall the assumption that $K(\cdot)$ is of second order, and that $f(u \mid \boldsymbol{z})$ has uniformly bounded derivatives with respect to u existing to

fourth order. These conditions yield

$$\frac{1}{h} E\left[K\left(\frac{u - \boldsymbol{x}^\top \boldsymbol{\beta}_0(U)}{h} \right) \right]$$

$$= \frac{1}{h} E\left[K\left(\frac{u - W_{\boldsymbol{x}}}{h} \right) \right]$$

$$= \frac{1}{h} \int K\left(\frac{u - w}{h} \right) f\left(w|\, \boldsymbol{z} \right) dw$$

$$= \int K(t) f\left(u - ht|\, \boldsymbol{z} \right) dt$$

$$= f\left(u|\, \boldsymbol{z} \right) + \frac{h^2}{2} f''\left(u|\, \boldsymbol{z} \right) \int s^2 K(s) ds + O\left(h^4 \right). \tag{31}$$

Combining (30) and (31) we get

$$\sup_{a_m \leq h \leq b_m} \sup_{\boldsymbol{z} \in \mathbb{R}^k} \sup_{\boldsymbol{x} \in \mathbb{R}^d} \sup_{\alpha \in \mathcal{A}} \left| \hat{f}_h\left(\boldsymbol{x}^\top \hat{\boldsymbol{\beta}}_n(\alpha) \middle| \boldsymbol{z} \right) - f\left(\boldsymbol{x}^\top \boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{z} \right) \right|$$

$$\leq \sup_{a_m \leq h \leq b_m} \sup_{\boldsymbol{z} \in \mathbb{R}^k} \sup_{\boldsymbol{x} \in \mathbb{R}^d} \sup_{\alpha \in \mathcal{A}} \left| \hat{f}_h\left(\boldsymbol{x}^\top \hat{\boldsymbol{\beta}}_n(\alpha) \middle| \boldsymbol{z} \right) - \frac{1}{h} E\left[K\left(\frac{\boldsymbol{x}^\top \hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{x}^\top \boldsymbol{\beta}_0(U)}{h} \right) \right] \right|$$

$$+ \sup_{a_m \leq h \leq b_m} \sup_{\boldsymbol{z} \in \mathbb{R}^k} \sup_{\boldsymbol{x} \in \mathbb{R}^d} \sup_{\alpha \in \mathcal{A}} \left| \frac{1}{h} E\left[K\left(\frac{\boldsymbol{x}^\top \hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{x}^\top \boldsymbol{\beta}_0(U)}{h} \right) \right] - f\left(\boldsymbol{x}^\top \boldsymbol{\beta}_0(U)|\, \boldsymbol{z} \right) \right|$$

$$\leq \sup_{a_m \leq h \leq b_m} \sup_{u \in \mathbb{R}} \left| \frac{1}{mh} \sum_{j=1}^{m} K\left(\frac{u - \boldsymbol{x}^\top \hat{\boldsymbol{\beta}}_n(U_j)}{h} \right) - \frac{1}{h} E\left[K\left(\frac{u - \boldsymbol{x}^\top \boldsymbol{\beta}_0(U)}{h} \right) \right] \right|$$

$$+ \sup_{a_m \leq h \leq b_m} \sup_{u \in \mathbb{R}} \left| \frac{1}{h} E\left[K\left(\frac{u - \boldsymbol{x}^\top \boldsymbol{\beta}_0(U)}{h} \right) \right] - f\left(u|\, \boldsymbol{z} \right) \right|$$

$$= O_p\left(\frac{1}{a_m^2 \sqrt{n}} + \sqrt{\frac{\log\left(\frac{1}{a_m} \right) \vee \log\log m}{m a_m}} \right) + O\left(b_m^2 \right)$$

as $m, n \to \infty$. $\qquad\qquad\square$

The conclusion of Lemma 4 is exploited to demonstrate the uniform convergence of the quantities $\hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha))$ and $\hat{\boldsymbol{\Delta}}_n^{-1}(\alpha)$ embedded in the expression for the test statistic $\hat{R}_n(g, \alpha)$ given in (8) above. In particular, we have the following:

Lemma 5. *Under the conditions of Assumptions 1–4, the following hold as $m, n \to \infty$:*

1. $\sup_{\alpha \in \mathcal{A}} \sup_{g \in \mathcal{G}} \left\| \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha)) - \boldsymbol{D}(g, \boldsymbol{\beta}_0(\alpha)) \right\| = o_p(1);$

2. $\sup_{\alpha} \left\| \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) - \boldsymbol{\Delta}^{-1}(\alpha) \right\| = o_p(1);$

where $\mathcal{A}$ is a compact subset of $[0, 1]$ and $\mathcal{G}$ is as described in Assumption 2.

45

Proof. We consider each part of the lemma in turn.

1. Consider $\left\| \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha)) - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))g(\boldsymbol{Z}_i) \right\|$. By the Cauchy-Schwarz inequality, we have

$$
\left\| \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha)) - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))g(\boldsymbol{Z}_i) \right\|^2
$$

$$
\leq \quad \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_{h_m}\left(\boldsymbol{X}_i^\top\hat{\boldsymbol{\beta}}_n(\alpha) \middle| \boldsymbol{Z}_i \right) - f\left(\boldsymbol{X}_i^\top\boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{Z}_i \right) \right)^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{X}_i\|^4
$$

$$
\cdot \frac{1}{n}\sum_{i=1}^{n}g^4(\boldsymbol{Z}_i)
$$

which is uniformly $o_p(1)$ as $m, n \to \infty$ for $\alpha \in \mathcal{A}$ and $g \in \mathcal{G}$ by Lemma 4 and Assumptions 1 and 3.

Now consider the asymptotic behaviour of

$$
\sup_{\alpha \in \mathcal{A}}\sup_{g \in \mathcal{G}}\left\| \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))g(\boldsymbol{Z}_i) - \boldsymbol{D}(g, \boldsymbol{\beta}_0(\alpha)) \right\| \tag{32}
$$

as $m, n \to \infty$. We argue that the quantity in (32) is $o_p(1)$ via an appeal to Lemma 1. In this connection, fix $\alpha_1 \in \mathcal{A}$ and $g_1 \in \mathcal{G}$. An application of the triangle inequality allows us to deduce that

$$
E\left[\sup_{\substack{\alpha: |\alpha - \alpha_1| < \delta, \\ g: \|g - g_1\|_{\mathcal{G}} < \delta}} \left\| f\left(\boldsymbol{X}_i^\top\boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{Z}_i \right)\boldsymbol{X}_i g(\boldsymbol{Z}_i) \right.\right.
$$

$$
\left.\left. - f\left(\boldsymbol{X}_i^\top\boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{Z}_i \right)\boldsymbol{X}_i g_1(\boldsymbol{Z}_i) \right\|^2 \right]
$$

$$
\leq \quad CE\left[|g_1(\boldsymbol{Z}_1)|^2 \sup_{\alpha: |\alpha - \alpha_1| < \delta}\left| f\left(\boldsymbol{X}_i^\top\boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{Z}_i \right) - f\left(\boldsymbol{X}_i^\top\boldsymbol{\beta}_0(\alpha_1) \middle| \boldsymbol{Z}_i \right) \right|^2 \cdot \|\boldsymbol{X}_i\|^2 \right]
$$

$$
+ CE\left[\sup_{g: \|g - g_1\|_{\mathcal{G}} < \delta} |g(\boldsymbol{Z}_i) - g_1(\boldsymbol{Z}_i)|^2 \right]
$$

$$
\leq \quad C\delta
$$

by Assumptions 1 and 3. Combining the conditions of Assumptions 1 and 3 with the compactness of $\mathcal{A}$, we have via Lemma 1 that the quantity in (32) is indeed $o_p(1)$ as $m, n \to \infty$ uniformly over $\alpha \in \mathcal{A}$ and $g \in \mathcal{G}$, as desired.

A final application of the triangle inequality yields the desired conclusion.

2. An appeal to Assumption 1 and Lemma 4 shows that

$$
\sup_{\alpha \in \mathcal{A}}\left\| \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) - \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))\boldsymbol{\delta}^\top(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \right)^{-1} \right\| = o_p(1)
$$

46

as $m, n \to \infty$. In addition,

$$\sup_{\alpha \in \mathcal{A}} \left\| \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \boldsymbol{\delta}^\top(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \right)^{-1} - \boldsymbol{\Delta}^{-1}(\alpha) \right\| = o_p(1)$$

as $n \to \infty$ by Assumption 1, the compactness of $\mathcal{A}$, a uniform law of large numbers and a continuous mapping theorem. Apply the triangle inequality and the desired conclusion is immediate.

$\square$

A.2 Proof of Theorem 1

We have

$$\hat{R}_n(g, \alpha)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n) \left(g(\boldsymbol{Z}_i) - \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha))^\top \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \right)$$

$$= S_n(g, \alpha) - \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha))^\top \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \hat{\boldsymbol{\beta}}_n) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha))$$

$$= S_n(g, \alpha) - \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha))^\top \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \cdot S_n \left(\hat{\boldsymbol{\delta}}_n(\cdot, \cdot, \hat{\boldsymbol{\beta}}_n(\alpha)), \alpha \right).$$

By Theorem 5 we have

$$\sup_{\alpha \in \mathcal{A}} \sup_{g \in \mathcal{G}} \left| S_n(g, \alpha) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i) - \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha)) \sqrt{n} \left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha) \right) \right|$$

$$= o_p(1). \tag{33}$$

Theorem 5 also allows us to write

$$\sup_{\alpha \in \mathcal{A}} \left| S_n \left(\hat{\boldsymbol{\delta}}_n \left(\cdot, \cdot, \hat{\boldsymbol{\beta}}_n(\alpha) \right), \alpha \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \hat{\boldsymbol{\delta}}_n(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha)) \right.$$

$$\left. - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(F \left(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha) \middle| \boldsymbol{Z}_i \right) - F \left(\boldsymbol{X}_i^\top \boldsymbol{\beta}_0(\alpha) \middle| \boldsymbol{Z}_i \right) \right) \cdot \hat{\boldsymbol{\delta}}_n \left(\boldsymbol{Z}_i, \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_n(\alpha) \right) \right|$$

$$= o_p(1),$$

which by Lemma 4 is equivalent to

$$\sup_{\alpha \in \mathcal{A}} \left| S_n \left(\hat{\boldsymbol{\delta}}_n \left(\cdot, \cdot, \hat{\boldsymbol{\beta}}_n(\alpha) \right), \alpha \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \right.$$

$$\left. - \boldsymbol{\Delta}(\alpha) \sqrt{n} \left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha) \right) \right|$$

$$= o_p(1). \tag{34}$$

47

Combining (33) and (34) we have that uniformly in $\alpha \in \mathcal{A}$ and $g \in \mathcal{G}$:

$$
\begin{aligned}
& \hat{R}_n(g, \alpha) \\
= \quad & \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i) + \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha)) \sqrt{n} \left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha) \right) \\
& - \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha))^\top \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \\
& - \hat{\boldsymbol{D}}_n(g, \hat{\boldsymbol{\beta}}_n(\alpha))^\top \hat{\boldsymbol{\Delta}}_n^{-1}(\alpha) \cdot \boldsymbol{\Delta}(\alpha) \cdot \sqrt{n} \left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha) \right) + o_p(1) \\
= \quad & \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \left(g(\boldsymbol{Z}_i) - \boldsymbol{D}(g, \boldsymbol{\beta}_0(\alpha))^\top \boldsymbol{\Delta}^{-1}(\alpha) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \right) + o_p(1) \\
= \quad & R_{n0}(g, \alpha) + o_p(1),
\end{aligned}
$$

where we have made appropriate use of Lemma 4.

A.3   Proof of Corollary 1

Note that

$$
R_{n0}(g, \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \left(g(\boldsymbol{Z}_i) - \boldsymbol{D}(g, \boldsymbol{\beta}_0(\alpha))^\top \boldsymbol{\Delta}^{-1}(\alpha) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha)) \right).
$$

The weak convergence of $R_{n0}(g, \alpha)$ follows from the joint weak convergence of

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i)
$$

and of

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))
$$

since $\boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha))$ is uniformly continuous in $\mathcal{G} \times \mathcal{A}$, as guaranteed by the conditions of Assumptions 1 and 3. The joint asymptotic equicontinuity follows from that of the marginals. A standard multivariate central limit theorem implies the convergence of the finite-dimensional distributions. To prove the asymptotic equicontinuity of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i)$, define for $\gamma \equiv (g, \alpha) \in \mathcal{G} \times \mathcal{A}$

$$
g_{1i}(\gamma) \equiv \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0(\alpha)) g(\boldsymbol{Z}_i)
$$

and define the class of functions $\mathcal{G}_1 = \{ g_{1\cdot}(\gamma) : \gamma \in \mathcal{G} \times \mathcal{A} \}$. Fix $\gamma_1 = (g_1, \alpha_1) \in \mathcal{G} \times \mathcal{A}$. By the triangle inequality, provided that $\delta \in (0, 1)$ we have

$$
\begin{aligned}
& E \left[\sup_{g_2 : \|g_1 - g_2\|_{\mathcal{G}} < \delta} \sup_{\alpha_2 : |\alpha_1 - \alpha_2| < \delta} |g_{1i}(\gamma_1) - g_{1i}(\gamma_2)|^2 \right] \\
\leq \quad & CE \left[\left\{ F \left(\boldsymbol{X}^\top \boldsymbol{\beta}_0(\alpha_1 + \delta) \big| \boldsymbol{Z} \right) - F \left(\boldsymbol{X}^\top \boldsymbol{\beta}_1(\alpha_1 - \delta) \big| \boldsymbol{Z} \right) \right\} \right] + C\delta \\
\leq \quad & C\delta.
\end{aligned}
$$

48

To prove the asymptotic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\alpha}(\boldsymbol{W}_i,\boldsymbol{\beta}_0)\boldsymbol{\delta}(\boldsymbol{Z}_i,\boldsymbol{X}_i,\boldsymbol{\beta}_0(\alpha))$, define for $\alpha \in \mathcal{A}$

$$g_{2i}(\alpha) = \psi_{\alpha}(\boldsymbol{W}_i,\boldsymbol{\beta}_0(\alpha))\boldsymbol{\delta}(\boldsymbol{Z}_i,\boldsymbol{X}_i,\boldsymbol{\beta}_0(\alpha)),$$

and define the class of functions $\mathcal{H}_1 = \{g_{2\cdot}(\alpha) : \alpha \in \mathcal{A}\}$. Fix $\alpha_1 \in \mathcal{A}$. By the triangle inequality, provided that $\delta \in (0,1)$, we have

$$
\begin{aligned}
&E\left[\sup_{\alpha_2\,:\,|\alpha_1-\alpha_2|<\delta}|g_{2i}(\alpha_1) - g_{2i}(\alpha_2)|^2\right] \\
\leq\ & CE\left[\{F\left(\boldsymbol{X}^\top\boldsymbol{\beta}_0(\alpha_1+\delta)\big|\,\boldsymbol{Z}\right) - F\left(\boldsymbol{X}^\top\boldsymbol{\beta}_1(\alpha_1-\delta)\big|\,\boldsymbol{Z}\right)\}\right] \\
&+ CE\left[\sup_{\alpha_2\,:\,|\alpha_1-\alpha_2|<\delta}\left|f\left(\boldsymbol{X}^\top\boldsymbol{\beta}_0(\alpha_1)\big|\,\boldsymbol{Z}\right) - f\left(\boldsymbol{X}^\top\boldsymbol{\beta}_0(\alpha_2)\big|\,\boldsymbol{Z}\right)\right|^2 \|\boldsymbol{X}\|^2\right] \\
\leq\ & C\delta.
\end{aligned}
$$

The desired conclusion follows.

A.4 Proof of Theorem 2

The proof of Theorem 2 parallels that of Theorem 1. In particular, we have

$$\hat{R}_n(g,\alpha) = S_n(g,\alpha) - \boldsymbol{D}^\top(g,\boldsymbol{\beta}_0(\alpha))\boldsymbol{\Delta}^{-1}(\alpha)S_n\left(\hat{\boldsymbol{\delta}}_n(\cdot,\cdot,\hat{\boldsymbol{\beta}}_n(\alpha)),\alpha\right) + o_p(1).$$

Applying Theorem 5 yields

$$
\begin{aligned}
&\sup_{\alpha\in\mathcal{A}}\sup_{g\in\mathcal{G}}\left|S_n(g,\alpha) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\alpha}(\boldsymbol{W}_i,\boldsymbol{\beta}_0)g(\boldsymbol{Z}_i) - E\left[a(\boldsymbol{Z},\boldsymbol{X},\alpha)g(\boldsymbol{Z})\right]\right. \\
&\left. - \boldsymbol{D}^\top(g,\boldsymbol{\beta}_0(\alpha))\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)\right)\right| \\
&= o_p(1),
\end{aligned}
$$

which by the orthogonality restriction $E\left[\boldsymbol{\delta}(\boldsymbol{Z},\boldsymbol{X},\boldsymbol{\beta}_0(\alpha))a(\boldsymbol{Z},\boldsymbol{X},\alpha)\right] = \boldsymbol{0}$ is equivalent to

$$
\begin{aligned}
&\sup_{\alpha\in\mathcal{A}}\sup_{g\in\mathcal{G}}\left|S_n\left(\hat{\boldsymbol{\delta}}_n\left(\cdot,\cdot,\hat{\boldsymbol{\beta}}_n(\alpha)\right),\alpha\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\alpha}(\boldsymbol{W}_i,\boldsymbol{\beta}_0)\boldsymbol{\delta}(\boldsymbol{Z}_i,\boldsymbol{X}_i,\boldsymbol{\beta}_0(\alpha))\right. \\
&\left. - \boldsymbol{\Delta}^{-1}(\alpha)\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)\right)\right| \\
&= o_p(1).
\end{aligned}
$$

As such, we have the following convergence holding uniformly for $g \in \mathcal{G}$ and $\alpha \in \mathcal{A}$ under the sequence of local alternatives given in (11):

$$
\hat{R}_n(g, \alpha)
$$
$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) g(\boldsymbol{Z}_i) + E\left[a(\boldsymbol{Z}, \boldsymbol{X}, \alpha) g(\boldsymbol{Z})\right] + \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0) \sqrt{n} \left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)\right)
$$
$$
- \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha)) \boldsymbol{\Delta}^{-1}(\alpha) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))
$$
$$
- \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha)) \boldsymbol{\Delta}^{-1}(\alpha) \cdot \boldsymbol{\Delta}(\alpha) \cdot \sqrt{n} \left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)\right) + o_p(1)
$$
$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(\boldsymbol{W}_i, \boldsymbol{\beta}_0) \left(g(\boldsymbol{Z}_i) - \boldsymbol{D}^\top(g, \boldsymbol{\beta}_0(\alpha)) \boldsymbol{\Delta}^{-1}(\alpha) \boldsymbol{\delta}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}_0(\alpha))\right)
$$
$$
+ E\left[a(\boldsymbol{Z}, \boldsymbol{X}, \alpha) g(\boldsymbol{Z})\right] + o_p(1),
$$

which completes the proof.

A.5 Proof of Theorem 3

The theorem follows from the multiplier central limit theorem; see van der Vaart and Wellner (1996, Theorem 2.9.2, p. 179).

Figure 1: Estimated cross-price elasticities of demand
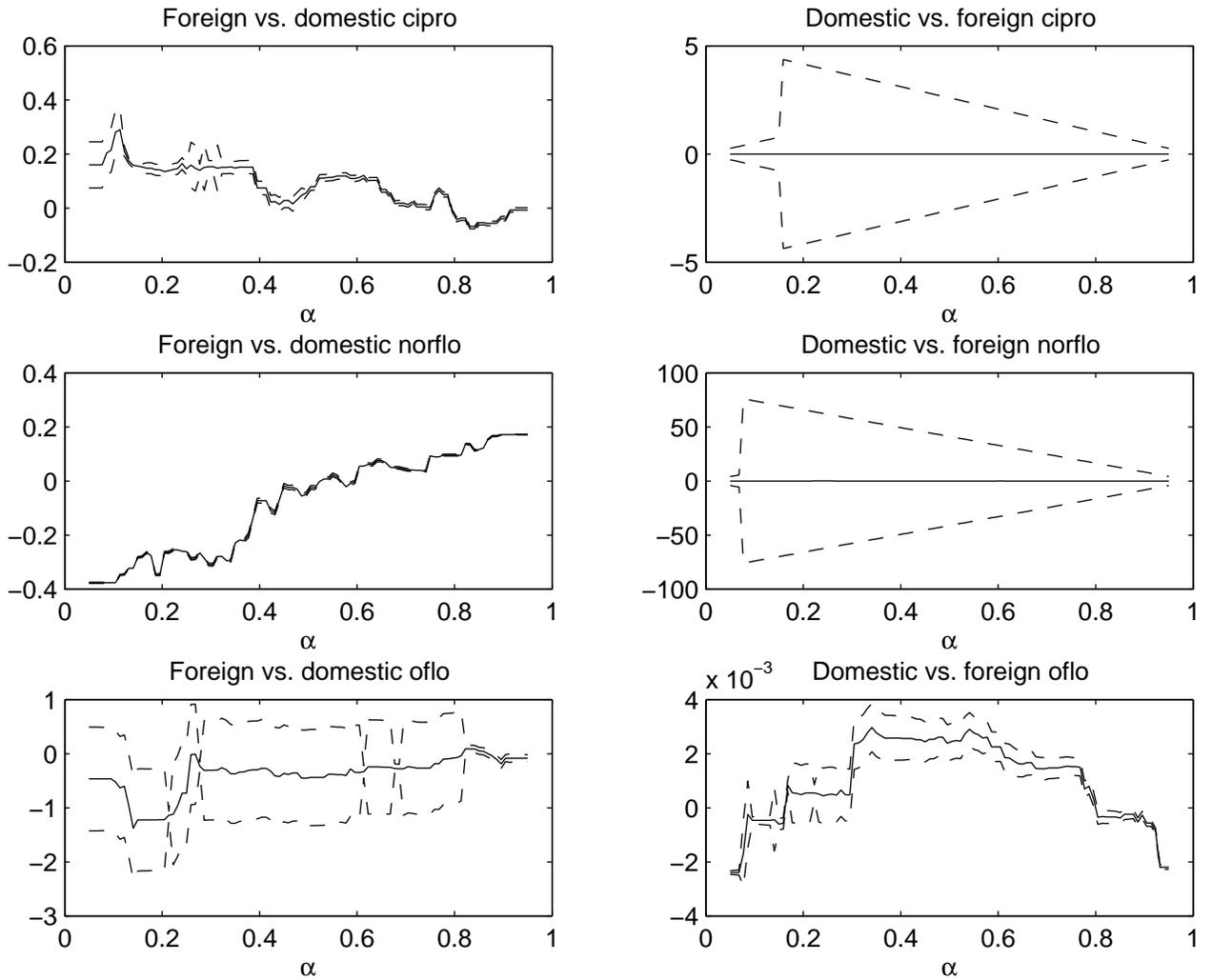
Note: The region between the dashed lines on each graph indicate 95-percent confidence regions.

51