Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 127 (2005) 225–252
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Unified approach to testing functional hypotheses in semiparametric contexts

Peter Hall[a,*], Adonis Yatchew[a,b]

[a]*Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia*
[b]*Department of Economics, University of Toronto, 150 St George St, Toronto, Ont. M5S 3G7, Canada 416-978-7128*

Received 4 February 2004
Available online 5 October 2004

Abstract

We suggest a new method, with very wide applicability, for testing semiparametric hypotheses about functions such as regression means and probability densities. The technique is based on characterising hypotheses in terms of functionals which can be estimated root-n consistently, and constructing test statistics in terms of estimators of the functionals. Since the tests are semiparametric it is appropriate to assess them on the basis of their ability to detect departures of size $n^{-1/2}$ from the null hypothesis. We show that they do indeed have this property. Unlike tests constructed in a nonparametric setting their power does not depend critically on choice of a bandwidth, and in particular, smoothing parameter selection is not an issue that has to be addressed by users of the tests. Bootstrap methods are suggested for calibrating the tests. In a regression setting, applications include tests of specification (such as partial linear and index models) against nonparametric or semiparametric alternatives, and tests of monotonicity, concavity, separability, equality of regression functions and base-independence of equivalence scales. In a density setting, they include tests of radial symmetry and stochastic dominance.

JEL classification: C14; D31

*Corresponding author.
E-mail addresses: halpstat@maths.anu.edu.au (P. Hall), yatchew@chass.utoronto.ca (A. Yatchew).

0304-4076/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2004.08.005

Keywords: Bootstrap; Convexity; Curve estimation; Density estimation; Independence; Index model; Monotonicity; Nonparametric; Parametric; Partial linear model; Positivity; Radial symmetry; Regression; Separability; Smoothing

1. Introduction

Rapidly increasing interest is being shown in hypothesis testing for infinite parameter problems, where the quantity under test is a function. The function is generally a regression mean or a probability density, although other contexts arise, including hazard rate functions. Sometimes the null hypothesis under test is qualitative in nature, for instance the assumption that the function is monotone or convex, or that one distribution stochastically dominates another, but in other cases it is more explicit. Examples in the latter context include index models, where it is argued that the true function might be of lower dimension than the data. Related hypotheses include those where it is supposed that the function is separable in some sense, for example as an additive model in regression, or as a multiplicative model in the case of a probability density.

Because such tests directly involve functions which, under at least the alternative hypothesis, are not known parametrically, then it is common to base them on relatively conventional nonparametric curve estimators, for example those constructed using kernel methods. While this approach can have advantages from some viewpoints, which we shall note two paragraphs below, in pragmatic terms it has two distinct disadvantages: first, it requires choice of a smoothing parameter, generally a bandwidth h; and secondly, when viewed as a test of semiparametric hypotheses, its power is unduly susceptible to the choice of h, often being particularly low if h is selected in a conventional manner.

In particular, if power is assessed in terms of ability to detect local semiparametric alternative hypotheses, then the nearest departures from the null hypothesis, H_0, that are detectable are generally $n^{-1/2}h^{-c}$ away, where n denotes sample size and $c>0$ depends on the context of the test. See, for example, Fan and Li (1996), Baltagi et al. (1996) and Lavergne and Vuong (2000), where $c = d/4$ (and d denotes dimension); and Anderson et al. (1998), where $c = d/2$. Since power obviously deteriorates with decreasing h, then, interpreting these quantities from a semiparametric viewpoint, it is advantageous to select h fixed. That choice, however, produces a test with asymptotically incorrect level, since the test is almost invariably calibrated on the basis of an argument (either asymptotic or bootstrap-based) which is invalid if the bandwidth does not converge to 0.

The alternative hypotheses that we consider are of the type $g = g_0 + n^{-1/2} g_1$, where g_0 is a fixed function satisfying H_0, g_1 is another fixed function, and $n^{-1/2} g_1$ denotes a perturbation which, when added to g_0, takes the latter out of the class H_0. In assessing the performance of tests in a nonparametric, rather than semiparametric, setting it is appropriate to take the alternative to be $g = g_0 + \delta_n g_n$, rather than $g_0 + n^{-1/2} g_1$, where g_n is a function chosen from a large class and whose

complexity may increase (usually through increasing frequency) with n. Again δ_n converges to zero as $n \to \infty$, but this time more slowly than $n^{-1/2}$. See Horowitz and Spokoiny (2001) for an example of such tests in econometrics.

An advantage of the nonparametric approach is its ability to capture departures $g = g_0 + \delta_n g_n$ from H_0, uniformly in a large class of g_n's. Its disadvantage, in a semiparametric rather than a nonparametric sense, is that it cannot detect relatively small departures of simpler type. There is no "free lunch"; the nonparametric test has "spent" a significant amount of sample information in order to develop great sensitivity to relatively complex, but more distant, departures from H_0, and that information cannot be effectively re-used to detect less sophisticated, but closer, departures.

In practice it can be quite difficult, particularly from the viewpoint of calibration, to identify the type of local departure from H_0 which is most plausible, and to construct a test accordingly. Doing so requires, in effect, combining a very large number of tests based, respectively, on a virtual continuum of different bandwidths. Instead, the test is usually constructed pragmatically from nonparametric curve estimators, using one bandwidth. This implicitly determines both g_n and δ_n, the latter as a function of the bandwidth, and that decision is generally made with only minor reference to the data. The method suggested by Horowitz and Spokoiny (2001) is an exception.

These difficulties constitute a major motivation for the semiparametric tests suggested in this paper. The tests address a very broad spectrum of hypotheses, and are able to detect semiparametric departures of as little as $n^{-1/2}$ from the null. None of the tests is based on curve or function estimation, and in particular none involves bandwidth choice, except at the calibration step (where bandwidth selection has negligible impact on power). To the contrary, the tests are founded on direct and explicit estimation of integrals of the unknown function over regions. Our estimators are computed directly from the data, rather than via smoothing methods. Reflecting this approach, the basis of our method is a new way of representing each member of a very large class of hypotheses about functions, in terms of properties of functionals of those functions. All the functionals are readily estimable root-n consistently.

A second motivation is the desirability of unified procedures for testing a variety of hypotheses in semiparametric settings. Apart from any conceptual and aesthetic appeal, one of the benefits of unification is that it facilitates testing of combinations of functional properties. For example, in the analysis of option pricing data, one might want to test whether the call option price is a monotone decreasing convex function of the strike price. Or, in examining demand data, one might want to simultaneously test homotheticity and consistency with the optimisation hypothesis.

A third motivation stems from the observation that economic models typically involve multiple explanatory variables. Even if the null is a highly structured semiparametric model, one cannot expect a great deal of power if the alternative is a pure nonparametric model of high dimension; see the third paragraph above. Thus, one would also like to test against plausible classes of semiparametric alternatives, which our procedure accommodates easily.

This is not to say that other testing methodologies cannot handle a broad range of hypotheses. Consider, for example, the residual regression approach. There, the idea is to estimate a model which incorporates the restrictions of the null, and then perform an "unrestricted" nonparametric regression of the estimated residuals on all explanatory variables to see whether anything remains to be explained. The idea has been used to construct tests of parametric or semiparametric (partial linear or index) specifications against nonparametric alternatives (Fan and Li, 1996; Zheng, 1996) and tests of equality of nonparametric regression functions in a panel data setting (Baltagi et al., 1996). The approach can readily be extended to test properties such as monotonicity, convexity or separability, and it easily accommodates common features of economic data such as heteroscedasticity. However, in addition to its dependence on bandwidth selection and lower power, it appears difficult to extend the procedure to settings where the alternative is semiparametric.

For the sake of brevity and simplicity, when developing our tests we shall focus on one-sample problems, where the hypotheses under test relate to the regression mean or density for the population from which a particular dataset is drawn. Multi-sample or multi-equation problems may be treated similarly, as we shall show in Section 6 where we shall treat tests of stochastic dominance and equality of nonparametric regressions. Our methods also give tests based on integral functions, sufficiently powerful to distinguish between the null hypothesis and an alternative distant $n^{-1/2}$ away. Indeed, early precursors of our method are the two-sample tests proposed by Hall and Hart (1990) and King et al. (1991). See also Hall et al. (1997), Koul and Schick (1997), Kulasekera and Wang (1997), Fan and Lin (1998) and Yatchew (1999). Reference to the tests proposed by Cramér and von Mises, and by Kolmogorov and Smirnov (see e.g. Kendall and Stuart, 1979, pp. 475–477), should also be made at this point.

Section 2 introduces our approach to characterising hypotheses in terms of functionals that can be estimated root-n consistently. There we quickly run through 10 examples, to give a flavour of the wide range of problems to which our methodology can be applied. Section 3 shows how to construct test statistics based on the proposals in Section 2, using the 10 examples for illustration. Section 4 outlines theoretical properties of the test statistic and suggests two methods for calibrating the test. The most attractive is arguably the bootstrap-based one; the other is founded on asymptotic properties of the test statistic under the null hypothesis. Section 5 describes power of the test against local alternatives, and gives simple, necessary and sufficient conditions for the test to be capable of detecting alternatives that are distant $n^{-1/2}$ from the null. The examples from Section 2 are used to show that these conditions hold very broadly. Numerical work, illustrating the performance of our approach, is summarised in Section 6. Finally, Section 7 outlines proofs of results in Section 5.

Reviews of testing methods include those of Hart (1997), Yatchew (1998) and Pagan and Ullah (1999). In the case of testing a parametric null hypothesis against a nonparametric or semiparametric alternative (Examples 1, 2 and 8 in Section 2), earlier contributors include Bierens (1982, 1990), Eubank and Spiegelman (1990), Eubank and Hart (1992), Wooldridge (1992), Azzalini and Bowman (1993),

Hong and White (1995), Bierens and Ploeberger (1997), Li and Wang (1998), Dette (1999) and Ellison and Ellison (2000). Work on testing for positivity, monotonicity and convexity (Examples 3 and 4 in Section 2) includes that of Schlee (1982), Yatchew (1992), Yatchew and Bos (1997), Bowman et al. (1998), Gijbels et al. (2000) and Hall and Heckman (2000). Tests for additive separability (Example 5) have been suggested by Barry (1993), Eubank et al. (1995) and Gozalo and Linton (2001). Fan and Li (1996) have proposed tests for semiparametric versus nonparametric alternatives (Examples 7 and 9). Horowitz and Härdle (1994) gave a test for a parametric or semiparametric null versus a semiparametric alternative (Example 10).

Performance criteria for nonparametric (as distinct from semiparametric) approaches to hypothesis testing have been discussed by, for example, Ingster (1982, 1993a, b), Härdle and Mammen (1993), Inglot et al. (1994, 2000), Spokoiny (1996) and Lepski and Spokoiny (1999). Horowitz and Spokoiny (2001) use the minimax approach introduced by Ingster as the basis for their approach.

2. Characterising a hypothesis by a functional

2.1. Nature of the characterisation

Let $\mathscr{G}$ denote a class of bounded functions g from one Euclidean space to another, equipped with the supremum metric. Write $\mathscr{G}_0$ for a particular subset of $\mathscr{G}$ that is of special interest. Given data from a model that features g, we wish to test a null hypothesis H_0 of the form $g \in \mathscr{G}_0$, against the alternative hypothesis H_1 that $g \in \mathscr{G} \backslash \mathscr{G}_0$.

If g was a regression mean then the data would typically be pairs of explanatory and response variables (X_i, Y_i), generated by the model

$$Y_i = g(X_i) + \varepsilon_i, \quad 1 \leqslant i \leqslant n, \tag{2.1}$$

where the ε_i's were independent errors with zero mean. The distribution of the pairs (X_i, ε_i) would generally not depend on choice of $g \in \mathscr{G}$. If g was a probability density then the data would usually be in the form of a random sample from the corresponding distribution.

Let Λ and Θ be sets, write $\mathscr{P}$ for the space of functions ψ from $\Lambda \times \Theta$ to the real line, and let $\psi(\cdot, \cdot \,|\, g)$ denote a particular element of $\mathscr{P}$, indexed by $g \in \mathscr{G}$. We want the functionals $\psi(\cdot, \cdot \,|\, g)$ to characterise $\mathscr{G}_0$, in the sense that

$$g \in \mathscr{G}_0 \quad \text{if and only if, for some } \theta \in \Theta, \psi(\lambda, \theta | g) = 0 \text{ for all } \lambda \in \Lambda. \tag{2.2}$$

Generally we choose λ to denote a region, or the finite vector of mathematical parameters that define a region, over which either g or some functional of it is integrated. The quantity θ represents a vector of statistical parameters. Usually the latter are the parameters of a conventional model, or of the parametric part of a semiparametric specification. In many instances, however, the role of θ is degenerate,

and there (2.2) is equivalent to

$$g \in \mathscr{G}_0 \quad \text{if and only if } \psi(\lambda|g) = 0 \text{ for all } \lambda \in \Lambda. \tag{2.3}$$

In a great many settings it is possible to choose the functionals $\psi(\cdot, \cdot \mid g)$ such that (a) characterisation (2.2) is achieved, and (b) each value of $\psi(\lambda, \theta|g)$ is estimable root-n consistently from data, in either the regression or the density estimation context. Property (b) is the key to constructing a powerful test based on the characterisation. Section 3 will show how to ensure it holds in a wide range of problems. In the remainder of the present section we shall outline 10 examples indicating the breadth of cases where (a) holds. Many more are possible, particularly in the contexts of tests involving probability densities and multi-sample or multi-equation problems.

2.2. Examples

In the first six examples Θ is degenerate. For simplicity, examples 1–4 and 8 are discussed in univariate settings, and there we take $\mathscr{G}$ to be a class of bounded and continuous functions from the interval $\mathscr{I} = [0, 1]$ to the real line. Nevertheless, each of these cases has a multivariate analogue which can be treated using our methods. Examples 5, 7, 9 and 10 are intrinsically multivariate.

In each example we give the set $\mathscr{G}_0$, which explicitly defines the null hypothesis H_0; see the first paragraph of Section 2.1. Then we show that either (2.3), or its more general form (2.2), holds. The set $\mathscr{G}$ of all possible candidates for g is usually the class of all functions that satisfy a smoothness condition, such as the existence of two bounded derivatives, and, in the case of hypothesis tests about densities, are probability density functions. However, in some instances $\mathscr{G}$ is restricted. Our methodology copes well with this case, as we show in Example 10.

In each example the definition of ψ that we give is chosen for its simplicity. There are many generalisations, perhaps the simplest of which is to include a weight function in the respective integral. This will alter properties of the test, and perhaps slightly improve power against some alternatives. However, optimal choice of the weight will depend on the particular alternative, and one usually has minimal information about that aspect of the problem.

Example 1 (*Equality to a specific function*). Put $\mathscr{G}_0 = \{g_0\}$, a singleton; the case $g_0 \equiv 0$ is of particular interest. Take Λ to be the set of intervals $\lambda = (\lambda_1, \lambda_2)$ with $0 < \lambda_1 < \lambda_2 < 1$, and put $\psi = \psi_1$ where

$$\psi_1(\lambda|g) = \int_\lambda (g - g_0)\,. \tag{2.4}$$

Then (2.3) holds. In a more general, multivariate setting, where each g might be a function from a subset $\mathscr{R}$ of d-variate Euclidean space to the real line, we could instead take Λ to be the set of all d-variate spheres, rectangles or similar scalable sets contained in $\mathscr{R}$, and again define $\psi(\lambda|g)$ as at (2.4).

Example 2 (*Linearity, and functions of specific type*). Let $\mathscr{G}_0$ be the set of all linear functions in $\mathscr{G}$. Take $\phi_0, \phi_1, \ldots$ to be any complete orthogonal sequence on the space of square-integrable functions from $\mathscr{I}$ to the real line, in which ϕ_0 and ϕ_1 are identically constant, and linear, respectively. (For example, $\{\phi_j\}$ might be the Jacobi sequence.) Let $\Lambda = \{2, 3, \ldots\}$, and put $\psi(\lambda|g) = \int_{\mathscr{I}} g\,\phi_\lambda$. Once again, (2.3) holds. See Example 8 for an alternative approach to testing this particular null hypothesis.

The example above has obvious extension to the problem of testing the hypothesis that g is a polynomial of specific degree, and to related settings.

Example 3 (*Monotonicity, convexity etc.*). Let $\mathscr{G}_0$ be the set of all nondecreasing functions in $\mathscr{G}$, let Λ be the set of all triples $(\lambda_1, \lambda_2, \lambda_3)$ where each component lies in $\mathscr{I}$ and $\lambda_2 - \lambda_1 = \lambda_3 - \lambda_2 > 0$, and put $\psi = \psi_2$ where

$$\psi_2(\lambda|g) = \min\left(\int_{\lambda_2}^{\lambda_3} g - \int_{\lambda_1}^{\lambda_2} g\,,\, 0\right). \tag{2.5}$$

Then (2.3) holds. The case where $\mathscr{G}_0$ is the set of all convex functions may be treated similarly, as too may the case where the constraint is that the jth derivative $g^{(j)}$ is everywhere of a given sign for a particular $j \geqslant 3$.

Example 4 (*Exceeding a given function*). This is related to Examples 1 and 3, and to testing for stochastic dominance; see Section 6 below. Write $\mathscr{G}_0$ for the set of all functions that never exceed a given function g_0:

$$\mathscr{G}_0 = \{g \in \mathscr{G} : g(x) \leqslant g_0(x) \text{ for all } x \in \mathscr{I}\}.$$

Here Λ may be taken to have the definition in Example 1, and $\psi = \max(\psi_1, 0)$ where ψ_1 is given by (2.4). Further examples, where $\mathscr{G}_0$ is the set of all functions for which $g^{(j)} - g_0^{(j)} \geqslant 0$ on $\mathscr{I}$, are similar. In particular, when $j = 1$ we may take Λ to be as in Example 3 and put $\psi(\lambda|g) = \psi_2(\lambda|g - g_0)$, where ψ_2 is given by (2.5).

Example 5 (*Separability*). In the bivariate case, $\mathscr{G}_0$ is the set of functions g that can be expressed as $g(u_1, u_2) = g_1(u_1) + g_2(u_2)$ for all (u_1, u_2) in a region $\mathscr{R}$. Now, $g \in \mathscr{G}_0$ if and only if, for any four points (x_1, x_2), (x_3, x_2), (x_1, x_4) and (x_3, x_4) forming a rectangle, we have

$$g(x_1, x_2) + g(x_3, x_4) - g(x_1, x_4) - g(x_3, x_2) = 0.$$

Hence, an equivalent condition is that, for any rectangle divided into four equal quadrants, the sum of the integrals for the SW and NE quadrants equals the sum for the NW and SE quadrants. Therefore we take Λ to be the set of all rectangles divided in this way, with their sides aligned to the coordinate axes. If $\lambda \in \Lambda$ has its corners at (x_1, x_2), (x_3, x_2), (x_1, x_4) and (x_3, x_4)

then we define

$$\begin{aligned}\psi(\lambda|g) = &\int_{x_2}^{(x_2+x_4)/2}\int_{x_1}^{(x_1+x_3)/2} g(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2\\ &+\int_{(x_2+x_4)/2}^{x_4}\int_{(x_1+x_3)/2}^{x_3} g(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2\\ &-\int_{x_2}^{(x_2+x_4)/2}\int_{(x_1+x_3)/2}^{x_3} g(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2\\ &-\int_{(x_2+x_4)/2}^{x_4}\int_{x_1}^{(x_1+x_3)/2} g(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2\,. \end{aligned} \tag{2.6}$$

Then (2.3) holds.

Higher-dimensional hypotheses of separability may be treated similarly. The problem of testing for independence in density estimation is related. There, in the bivariate case, one would take Λ to be the set of discs, rectangles or similar scalable sets in the plane, and define

$$\psi(\lambda|g) = \int_\lambda \{g(x_1,x_2) - g_1(x_1)\,g_2(x_2)\}\,\mathrm{d}x_1\,\mathrm{d}x_2\,, \tag{2.7}$$

where $g_1(x_1) = \int g(x_1,x_2)\,\mathrm{d}x_2$ and g_2 is defined analogously.

Example 6 (*Radial symmetry*). This property is of interest in density estimation for d-variate distributions, partly because it mitigates the "curse of dimensionality". Under radial symmetry, $\mathscr{G}_0$ is the set of densities g such that $g(x)$ varies only with the distance of x from a given point. That point is generally known, and so we may take it to be the origin. Without loss of generality g is expressed in polar coordinates, as $g(r,\omega)$ where r denotes distance from the origin and ω is a $(d-1)$-variate vector of angles. Take $\mathscr{O}$ to be the set of possible values of ω, let Λ be the set of cross products of members of a class of scalable subsets of $\mathscr{O}$, such as rectangles, with compact subintervals of $(0,\infty)$, representing radius; and put

$$g_1(r) = \frac{\int_{\mathscr{O}} g(r,\omega)\,\mathrm{d}\omega}{\int_{\mathscr{O}}\mathrm{d}\omega}$$

and

$$\psi(\lambda|g) = \int_\lambda \{g(r,\omega) - g_1(r)\}\,\mathrm{d}r\,\mathrm{d}\omega\,. \tag{2.8}$$

Then (2.3) holds. Elliptical symmetry may be accommodated by a suitable change in coordinate system.

Example 7 (*Partial linear model*). In a bivariate setting where the potential linear relationship applies to the second component, $\mathscr{G}_0$ is the set of all functions g that can be represented as $g(x_1,x_2) = g_1(x_1) + \theta x_2$ for a scalar parameter θ and all $(x_1,x_2) \in \mathscr{R}$. Take Λ to be the set of all discs (or rectangles or similar bivariate scalable sets)

contained in $\mathscr{R}$, define $\mathscr{R}(x_1)$ to be the set of all x_2 such that $(x_1, x_2) \in \mathscr{R}$, and put

$$g(x_1|\theta) = \frac{\int_{\mathscr{R}(x_1)} \{g(x_1, x_2) - \theta x_2\}\, \mathrm{d}x_2}{\int_{\mathscr{R}(x_1)} \mathrm{d}x_2}$$

and

$$\psi(\lambda, \theta|g) = \int_\lambda \{g(x_1, x_2) - g(x_1|\theta) - \theta x_2\}\, \mathrm{d}x_1\, \mathrm{d}x_2\,. \tag{2.9}$$

Then (2.2) holds. Higher-dimensional problems admit similar treatment.

A broad range of examples where Θ plays an important role arise in the context of so-called specification tests, where parametric or semiparametric null hypotheses are tested against semiparametric or nonparametric alternatives. Our remaining three examples treat such cases.

Example 8 (*Parametric null hypothesis*). Consider the parametric setting, where $\mathscr{G}_0 = \{g_1(\cdot|\theta) : \theta \in \Theta\}$, Θ is a class of parameters, and for each $\theta \in \Theta$, $g_1(\cdot|\theta)$ is a continuous, known function on $\mathscr{I}$. Take Λ to be as in Example 1 and, analogously to (2.4), put

$$\psi(\lambda, \theta|g) = \int_\lambda \{g - g_1(\cdot|\theta)\}\,.$$

Then (2.2) holds. This example can be extended to the multivariate case by arguing as in the remark at the end of Example 1.

Example 9 (*Multiple index model*). Assume $\mathscr{G}$ is a class of smooth, bounded functions from a d-variate Euclidean space to the real line. Given $1 \leqslant r < d$, let $\mathscr{T}_r$ denote the set of all $r \times d$ matrices θ with orthonormal rows. We call $g \in \mathscr{G}$ an r-variate index model if, for some $\theta \in \mathscr{T}_r$ and some function γ of r variables, $g(x) = \gamma(\theta x)$ for each x. Taking $r = 1$ gives the common "single index model". For $r \geqslant 2$, a rule for ordering the rows of θ removes redundancy created by permuting the rows. We shall assume this has been done.

Write $\mathscr{G}_0$ for the set of all r-variate index models. In the case where the domain of g is a d-variate sphere (without loss of generality, the d-variate unit sphere $\mathscr{S}_d$, centred at the origin), we shall suggest a class of functionals $\psi(\cdot, \cdot \,|\, g)$ for which (2.2) holds and Λ does not depend on θ. More generally, when the domain is an asymmetric region, Λ must be permitted to vary with θ.

Let θ_{ident} be the "identity" version of θ, where the component in position (i, i) equals 1 for $1 \leqslant i \leqslant r$ and each other component is 0. Let Λ denote the class of "cylinders" representable as the product of an r-variate sphere with a $(d - r)$-variate rectangular prism, the former in an r-dimensional plane parallel to the one having its axes in the directions of the unit vectors comprising the rows of θ_{ident}, and wholly contained within $\mathscr{S}_d$. Given $\lambda \in \Lambda$ and $\theta \in \Theta$, let λ_θ denote the result of rotating λ about its centre so that its spherical base is parallel to the plane defined by θ rather than θ_{ident}. Let $\tau(\lambda, \theta)$ denote the cylinder that has the same axis of symmetry as λ_θ but is as long as possible (in the direction of this axis) subject to being wholly

contained in $\mathscr{S}_d$. Put

$$\psi(\lambda,\theta|g) = \int_{\lambda_\theta} g(x)\,\mathrm{d}x - \frac{\int_{\lambda_\theta}\mathrm{d}x}{\int_{\tau(\lambda,\theta)}\mathrm{d}x}\int_{\tau(\lambda,\theta)} g(x)\,\mathrm{d}x\,. \tag{2.10}$$

Then (2.2) holds.

Example 10 (*Linear versus linear index model*). In the bivariate case, $\mathscr{G}_0$ is the class of functions of the form $g(x_1,x_2)=\alpha_1+\alpha_2x_1+\alpha_3x_2$ for scalars $\alpha_1,\alpha_2,\alpha_3$, and $\mathscr{G}$ is the set of g of the form $g(x_1,x_2)=g_1(x_1+\alpha x_2)$ for a scalar α and a univariate function g_1 determined only semiparametrically. Here we take Λ to be the set of discs, rectangles or similar scalable set in the plane, and, using a slight reparametrisation, put

$$\psi(\lambda,\theta|g) = \int_{\lambda_\theta} [\,g_1(x_1+\theta_0x_2) - \{\theta_1+\theta_2\,(x_1+\theta_0x_2)\}\,]f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\,, \tag{2.11}$$

where f is the design density. Then (2.2) holds. Higher-dimensional problems may be treated similarly.

Tests of various other properties, and combinations thereof, can be constructed, in some cases by simple adaptation of the above examples. To construct a test of the relevance or significance of a variable in a regression model, e.g. a test of whether the function $g(x_1,x_2)$ is constant with respect to x_2, set $\theta=0$ in Example 7. Note that by suitable definition of variables, homogeneity of degree zero and homotheticity may be tested using such a test of significance.

3. Constructing the test statistic

3.1. *Encapsulating* H_0 *in a single integral*

Except for the case of Example 2 in Section 2.2, a key property that all the functionals ψ in our examples enjoy is that they are continuous in their argument λ. In particular,

> If, for some g and θ, $\psi(\lambda,\theta|g)$ is nonzero for $\lambda=\lambda_0$, say,
> then it is nonzero for all λ in some neighbourhood of λ_0. (3.1)

This means that we can easily encapsulate characterisations such as (2.2) and (2.3) in terms of integral functions of the functional $\psi(\,\cdot,\cdot\,|\,g)$.

This may be done in a variety of ways; we shall consider only one. Let A denote a continuous real-valued function of a real variable, vanishing only at the origin and strictly positive elsewhere. Let μ denote a bounded, continuous, strictly positive measure on Λ. Given $\psi\in\mathscr{P}$ (the latter defined in Section 2.1), put

$$\langle\psi\rangle_\theta = \int A\{\psi(\lambda,\theta)\}\,\mu(\mathrm{d}\lambda)\,.$$

In view of property (3.1), and the fact that μ is strictly positive on Λ, characterisation (2.2) is equivalent to: $g \in \mathscr{G}_0$ if and only if $\inf_\theta \langle \psi(\cdot, \cdot \mid g) \rangle_\theta = 0$. In those instances where the role of θ is degenerate (see Examples 1–6 in Section 2.2) we may drop the infimum over θ, in which case (2.2) is equivalent to: $g \in \mathscr{G}_0$ if and only if $\int A\{\psi(\cdot|g)\}\, \mu(\mathrm{d}\lambda) = 0$.

For simplicity and definiteness, in the remainder of this paper we shall take $A(u) \equiv u^2$ and work with the square root of the corresponding criterion $\langle \psi \rangle$. Other choices of A give statistics with different asymptotic distributions and slightly different power properties, although their ability to distinguish alternatives that are distant $n^{-1/2}$ from the null hypothesis remains unaltered.

Thus, we put

$$\|\psi\|_\theta = \left\{ \int \psi(\lambda, \theta)^2 \, \mu(\mathrm{d}\lambda) \right\}^{1/2} \tag{3.2}$$

and $\|\psi\| = \sup_\theta \|\psi\|_\theta$. Then $\|\cdot\|$ is a norm on $\mathscr{P}$. Defining

$$t(g) = \inf_\theta \|\psi(\cdot, \cdot \mid g)\|_\theta \tag{3.3}$$

we see that, in view of (3.1) and the fact that μ is positive on Λ,

$$g \in \mathscr{G}_0 \quad \text{if and only if } t(g) = 0\,. \tag{3.4}$$

Therefore we may assess the veracity of H_0 by estimating $t(g)$ and testing the significance of its difference from zero. If the role of θ is degenerate then we define $t(g) = \|\psi(\cdot|g)\|$, where $\|\psi\|^2 = \int \psi(\lambda)^2 \, \mu(\mathrm{d}\lambda)$. Result (3.4) continues to hold.

In all but the second of the 10 examples described in Section 2.2, the sets Λ may be considered to be indexed by a finite number of discrete parameters which vary in the continuum. For example, in the case of Example 1 these are the pairs (λ_1, λ_2) with $0 < \lambda_1 < \lambda_2 < 1$; for Example 3 they are triples $(\lambda_1, \lambda_2, \lambda_3)$ with $0 < \lambda_1 < \lambda_2 < \lambda_3 < 1$ and $\lambda_3 - \lambda_2 = \lambda_2 - \lambda_1$; and so on. We should take μ to be an absolutely continuous and strictly positive measure on the class of parameters that determine Λ; see Section 3.4 for examples. Asymptotic theory is more straightforward if the measure μ is bounded, but that may not be necessary in practice.

In the contrary case of Example 2, where Λ is a countable discrete set, Property (3.1) is not well defined. There, however, if we take μ to be a bounded measure which employs a strictly positive weight for each element of Λ, then (3.4) follows directly from (2.1).

3.2. Test statistic

Assume we can construct a root-n consistent estimator $\widehat{\psi}(\cdot, \cdot \mid g)$ of $\psi(\cdot, \cdot \mid g)$, using either the data at (2.1), in the event that g is a regression mean, or a random sample from a distribution with density g, if the hypotheses concern densities. Our test statistic is the following estimator of $t(g)$:

$$T = T(g) = \inf_\theta \|\widehat{\psi}(\cdot, \cdot \mid g)\|_\theta\,. \tag{3.5}$$

We reject H_0 in favour of H_1 if T is too large, where the meaning of "too large" could be determined using bootstrap methods for calibration.

In cases where dependence of ψ on θ is degenerate, $\widehat{\psi}(\lambda|g)$ is an estimator of $\psi(\lambda|g)$ and our test statistic is simply

$$T = \left\{ \int \widehat{\psi}(\lambda|g)^2 \, \mu(\mathrm{d}\lambda) \right\}^{1/2}. \tag{3.6}$$

For the sake of brevity, when considering ways of estimating ψ we shall for the most part consider only the regression case, where data (X_i, Y_i) are generated by the model at (2.1). The setting where g is a probability density is similar, and in fact is a little simpler since one does not need to account for the distribution of the design points X_i. We consider that aspect next.

3.3. Correcting for nonuniform design

We shall assume that the X_i's are independent and identically distributed with density f; they may be multivariate. If f is constant, i.e. if the X_i's are uniformly distributed, then the functions ψ in the examples in Section 2.2 can easily be estimated by taking sums of the response variables Y_i over indices i corresponding to the X_i's lying in particular sets. When f is not constant, however, we need to correct by multiplying by an approximation to $1/f$.

This is a familiar issue in the context of nonparametric regression. There, the Nadaraya–Watson estimator solves the problem by explicitly estimating $1/f$; see for example p. 130 of Wand and Jones (1995). We could adopt that approach here, but it requires bandwidth choice and, furthermore, is not in the spirit of our work. Moreover, we do not require a consistent estimator of f; a stochastically varying estimator, for which the stochastic errors were approximately equally distributed and therefore virtually cancelled one another in the sums that we shall use to construct estimators of ψ functions, would be adequate.

With this in mind we suggest the following near-neighbour approximation to $1/f$; it does not require selection of a smoothing parameter. Assume the design variables are d-variate, write $D_j(x)$ for the distance from x to the jth nearest X_i that does not equal x, and let D be a linear combination of the functions D_j^d, chosen so that $nv_d \, \mathrm{E}\{D(x)\} \approx 1/f(x)$, where $v_d = \pi^{d/2}/\Gamma(1 + \frac{1}{2}d)$ is the content of a d-variate sphere of unit radius. Then, provided we multiply Y_i by $W_i \equiv nv_d \, D(X_i)$ whenever we include the former in a series approximation to the integral ψ, we correct for the design density.

We should construct our approximation to $1/f$ so that it does not introduce bias terms to the asymptotic distribution of our root-n consistent estimator $\widehat{\psi}$. Such terms are intrinsically difficult to accommodate using bootstrap methods. It is readily seen that this requires

$$nv_d \, \mathrm{E}\{D(x)\} = f(x)^{-1} + \mathrm{o}(n^{-1/2}). \tag{3.7}$$

Assuming that f is bounded away from zero in a neighbourhood of x and has two bounded derivatives there, (3.7) holds for $d = 1, 2, 3$ if we take simply $D = D_1^d$. Provided f has four bounded derivatives in the neighbourhood, and $1 \leqslant d \leqslant 7$, it suffices to take

$$D = \frac{\Gamma(3 + 2d^{-1})\, D_1^d - \Gamma(2 + 2d^{-1})\, D_2^d}{\Gamma(3 + 2d^{-1}) - 2\,\Gamma(2 + 2d^{-1})}\,,$$

where Γ denotes the standard gamma function. Similar approximations are readily obtained for larger values of d.

Standard arguments, based on the fact that $D_j(x)$ has a Binomial distribution, show that in either of these cases, and to first order, W_i equals $Z_i/f(X_i)$ plus smaller order terms, where Z_i has an exponential distribution with unit mean and is independent of X_i. The effect of Z_i must of course be incorporated into approximations to the distribution of the test statistic T.

In some cases the nonuniformity of design density does not matter. Instances in point are those of Examples 1, 4, 8 and 10, where a simple modification of ψ to incorporate f overcomes difficulties. In particular, if in Example 1 we alter the definition of $\psi = \psi_1$ at (2.4) to

$$\psi(\lambda|g) = \int_\lambda (g - g_0) f\,, \tag{3.8}$$

then no correction for f is required. We can estimate this version of $\psi(\lambda|g)$ directly, using

$$\widehat{\psi}(\lambda|g) = n^{-1} \sum_{i\,:\,X_i \in \lambda} \{Y_i - g_0(X_i)\}\,. \tag{3.9}$$

(Recall that Example 1 is univariate, and that λ is a subinterval of the support interval $\mathscr{I}$ of the design density.) This estimator $\widehat{\psi}(\lambda|g)$ is unbiased, as well as root-n consistent, for $\psi(\lambda|g)$. Provided f is continuous and does not vanish on $\mathscr{I}$, the functionals $\psi(\cdot|g)$ at (3.8) characterise the null hypothesis in the sense of (2.3): $g = g_0$ if and only if $\psi(\lambda|g) = 0$ for all $\lambda \in \Lambda$. Examples 4, 8 and 10 admit similar treatments, as too do their multivariate counterparts.

3.4. Constructing the estimators $\widehat{\psi}(\cdot, \cdot\,|\,g)$

We shall briefly run through the 10 examples given in Section 2.2, showing how in each case an estimator $\widehat{\psi}(\cdot, \cdot\,|\,g)$ of $\psi(\cdot, \cdot\,|\,g)$ can be constructed. In Section 4.1 we shall refer back to these examples when outlining the form taken by asymptotic theory for the statistic T at (3.6). It will be clear from our constructions that the estimators $\widehat{\psi}$ are root-n consistent; they are means of order n independent random variables. In Section 4.1 we shall note that the estimators usually also satisfy central limit theorems; see formula (4.1).

At (3.9) we suggested one means of accommodating nonuniform design density in the context of Example 1. Another approach is to employ the methods suggested in

the previous section, using

$$\widehat{\psi}(\lambda|g) = n^{-1} \sum_{i\,:\,X_i \in \lambda} \{Y_i - g_0(X_i)\}\, W_i$$

as an estimator of $\psi(\lambda|g) = \int_\lambda (g - g_0)$, rather than using the statistic at (3.9) to estimate the quantity at (3.8).

For Example 2 we take $\widehat{\psi}(\lambda|g) = n^{-1} \sum_i Y_i\, \phi_\lambda(X_i)\, W_i$. In the case of Example 3 we employ

$$\widehat{\psi}(\lambda|g) = \min\left(n^{-1} \sum_{i\,:\,\lambda_2 < X_i < \lambda_3} Y_i\, W_i - n^{-1} \sum_{i\,:\,\lambda_1 < X_i < \lambda_2} Y_i\, W_i\,,\, 0 \right)$$

as an estimator of $\psi = \psi_2$ defined at (2.5). Example 4 is similar to Example 1.

In Example 5, we estimate $\psi(\lambda|g)$ at (2.6) using $\widehat{\psi}(\lambda|g) = S_1 + S_2 - S_3 - S_4$, where $S_1, \ldots, S_4$ denote the sums of $Y_i\, W_i$ over indices i such that X_i lies in the rectangular regions over which the four respective integrals at (2.6) are taken. In the related problem of testing independence of marginals of a probability density, where data $X_i = (X_{1i}, X_{2i})$, $1 \leqslant i \leqslant n$, are drawn from a bivariate distribution with density g, and where $\psi(\lambda|g)$ is given by (2.7) for rectangles $\lambda = \lambda_1 \times \lambda_2$, we estimate it by

$$\widehat{\psi}(\lambda|g) = n^{-1} \sum_{i=1}^{n} I(X_i \in \lambda) - \left\{ n^{-1} \sum_{i=1}^{n} I(X_{1i} \in \lambda_1) \right\} \left\{ n^{-1} \sum_{i=1}^{n} I(X_{2i} \in \lambda_2) \right\}.$$

Example 6 also involves density estimation. There, if data $X_i = (R_i, \Omega_i)$, $1 \leqslant i \leqslant n$, are recorded in polar coordinates, and if Λ is a class of sets $\lambda = \lambda_1 \times \lambda_2$ where λ_1 is a subset of the positive real line and λ_2 is a subset of $\mathcal{O}$, then the function $\psi(\lambda|g)$ defined at (2.8) is estimated by

$$\widehat{\psi}(\lambda|g) = n^{-1} \sum_{i=1}^{n} I(X_i \in \lambda) - \left\{ n^{-1} \sum_{i=1}^{n} I(R_i \in \lambda_1) \right\} \frac{\int_{\lambda_2} \mathrm{d}\omega}{\int_{\mathcal{O}} \mathrm{d}\omega}\,.$$

In the context of Example 7, let Λ denote the set of rectangles λ that have their axes aligned with those of the coordinate system for the design points $X_i = (X_{1i}, X_{2i})$, write $\lambda = \lambda_1 \times \lambda_2$ where λ_1 and λ_2 are intervals, let f_1 be the marginal density of X_{1i}, let W_{1i} denote the version of W_i for approximating $f_1(X_{1i})$ rather than $f(X_i)$, and as our estimator of $\psi(\lambda, \theta|g)$ at (2.9), take

$$\widehat{\psi}(\lambda, \theta|g) = n^{-1} \sum_{i\,:\,X_i \in \lambda} Y_i\, W_i - \frac{|\lambda_2|}{n} \sum_{i\,:\,X_{1i} \in \lambda_1} (Y_i - \theta X_{2i})\, W_{1i} - \theta \int_\lambda x_2\, \mathrm{d}x_1\, \mathrm{d}x_2\,, \tag{3.10}$$

where $|\lambda_2|$ denotes the length of the interval λ_2.

Example 8 is similar to Example 1. For Example 9 we estimate $\psi(\lambda, \theta|g)$ at (2.10) by

$$\widehat{\psi}(\lambda, \theta|g) = n^{-1} \sum_{i\,:\,X_i \in \lambda_\theta} Y_i\, W_i - \frac{\int_{\lambda_\theta} \mathrm{d}x}{n \int_{\tau(\lambda,\theta)} \mathrm{d}v} \sum_{i\,:\,X_i \in \tau(\lambda,\theta)} Y_i\, W_i. \tag{3.11}$$

In the case of Example 10 we estimate $\psi(\lambda, \theta|g)$ at (2.11) by

$$\widehat{\psi}(\lambda, \theta|g) = n^{-1} \sum_{i\,:\,X_{1i} + \theta_0 X_{2i} \in \lambda_\theta} [Y_i - \{\theta_1 + \theta_2 (X_{1i} + \theta_0 X_{2i})\}].$$

In some of these examples the estimator $\widehat{\psi}$ involves ratios of random variables; see (3.10) and (3.11). From at least a practical viewpoint the denominators of these ratios must be kept reasonably large, to prevent unduly large stochastic fluctuations. This means that the regions λ should not be taken too small.

4. Calibration

4.1. Asymptotic properties of T

Calibration can be based on the large-sample distribution of T. However, while this is feasible it is unattractive, since the limiting distribution is a complicated function of a Gaussian process the properties of which depend on unknowns. A bootstrap approach, outlined in Section 4.2, is more attractive. Nevertheless, large-sample properties of T are important to understanding their analogues for the bootstrap version of T, so we describe them here.

In all our examples, and in all circumstances where we anticipate our tests being applied, Λ is either a set of vectors of given finite length, or a set of regions determined by such vectors. Without loss of generality, the former is true. It is generally the case that, under the null hypothesis, the estimator $\widehat{\psi}(\lambda, \theta|g)$ of $\psi(\lambda, \theta|g)$ converges weakly in the conventional sense of convergence of stochastic processes, indexed by the finite vector (λ, θ). That is, for each $g \in \mathscr{G}_0$ there exists a Gaussian process $\zeta(\cdot, \cdot\,|\,g)$ with zero mean, defined on $\Lambda \times \Theta$, such that

$$n^{1/2}\,\{\widehat{\psi}(\cdot, \cdot\,|\,g) - \psi(\cdot, \cdot\,|\,g)\} \to \zeta(\cdot, \cdot\,|\,g) \tag{4.1}$$

weakly on $\Lambda \times \Theta$. The limiting process $\zeta(\cdot, \cdot\,|\,g)$ has continuous sample paths, and the convergence in (4.1) is in the sense of the uniform topology. The next paragraphs but one discuss theoretical arguments.

Cases where ψ is defined in terms of a minimum, such as those arising in Examples 3 and 4, should be treated slightly differently. There (4.1) does not hold, although it is nevertheless true that

$$\psi = \min(\chi, 0) \text{ and } \widehat{\psi} = \min(\widehat{\chi}, 0), \text{ where } n^{1/2}(\widehat{\chi} - \chi) \to \zeta \text{ in distribution}. \tag{4.2}$$

Some cases of (4.1) are trivial to prove. For instance, if $\widehat{\psi}(\lambda|g)$ at (3.9) is used to estimate $\psi(\lambda|g)$ at (3.8), if we assume regression model (2.1) in which the errors ε_i are independent and identically distributed random variables with zero mean and finite variance, and if the X_i's are independent and identically distributed with a continuous, compactly supported density f, then (4.1) follows from conventional invariance principles for sums of independent random variables. The case where $\varepsilon_i = \sigma(X_i)\,\delta_i$, for a bounded, continuous, positive function σ and independent and identically distributed variables δ_i, is similar.

We claim that, under the null hypothesis,

$$n^{1/2}\,T \to U \tag{4.3}$$

in distribution, where the random variable U has a continuous distribution and may be expressed as a functional of the stochastic process ζ in (4.1). Of all our 10 examples the case of multiple-index models (Example 9) is the most awkward to treat. A detailed and rigorous derivation, in the case of single-index models, of (4.1), (4.3), and its bootstrap analogue (4.5), is given in the single index case by Delecroix et al. (2002), and in the Université de Rennes doctoral thesis of Céline Vial. In particular this work describes theory for the density corrections W_i. Theory in the multiple-index case is similar. While the other nine examples are simpler, it does not seem possible to treat all examples together in a reasonable amount of space. Therefore, here and in Section 4.2 we shall confine ourselves to outlining the general argument.

Note first that when the role of θ is degenerate, as in Examples 1–6, (4.1) simplifies to $n^{1/2}\widehat{\psi}(\cdot|g) \to \zeta(\cdot|g)$ under H_0. (The arguments here are simply λ.) Hence, $n^{1/2}\,T \to U$ in distribution, where $U^2 = \int \zeta(\lambda|g)^2\,\mathrm{d}\mu(\lambda)$. When the role of θ is not degenerate we may approximate the distribution of $\|\widehat{\psi}(\,\cdot\,,\cdot\,|\,g)\|_\theta^2$ by that of

$$\begin{aligned}&\int \{\psi(\lambda,\theta_0) + (\theta-\theta_0)^{\mathrm{T}}\nabla\psi(\lambda,\theta_0) + n^{-1/2}\zeta(\lambda,\theta_0|g)\}^2\,\mu(\mathrm{d}\lambda)\\ &\quad = \int \{(\theta-\theta_0)^{\mathrm{T}}\nabla\psi(\lambda,\theta_0) + n^{-1/2}\zeta(\lambda,\theta_0|g)\}^2\,\mu(\mathrm{d}\lambda)\,,\end{aligned}$$

where $\theta_0 = \mathrm{argmin}_\theta\,\|\psi(\,\cdot\,,\cdot\,|\,g)\|_\theta$ and $\nabla\psi(\lambda,\theta|g)$ is the vector of partial derivatives of $\psi(\lambda,\theta|g)$ with respect to θ. It follows that $n^{1/2}\,T \to U$ in distribution, where, assuming θ is a vector of length k,

$$U^2 = \inf_{\beta\in\mathbb{R}^k} \int \{\beta^{\mathrm{T}}\nabla\psi(\lambda,\theta_0) + \zeta(\lambda,\theta_0|g)\}^2\,\mu(\mathrm{d}\lambda)\,. \tag{4.4}$$

Again the case where ψ is defined in terms of a minimum, as in Examples 3 and 4, needs a slightly different treatment. It can be deduced from (4.2), however, that the result that $n^{1/2}T$ converges in distribution to a random variable U remains true.

4.2. Bootstrap calibration

For brevity we shall again confine attention to the setting of regression. The case of density estimation is simpler. In regression the bootstrap calibration step requires

two estimators of g. One, usually a standard nonparametric smoother to which we shall refer as $\tilde{g}$, is used only to compute residuals (i.e. estimators of the errors ε_i). Therefore, it does not need to be particularly accurate; in asymptotic terms it is required only to be uniformly consistent for the true g. However, it should have this property regardless of the validity of the null hypothesis. The second estimator, to which we shall refer as $\hat{g}$, should be an element of the class $\mathscr{G}_0$ that determines the null hypothesis. It should be uniformly consistent for g if the null hypothesis is correct, i.e. if $g \in \mathscr{G}_0$, but of course it will be inconsistent otherwise. (In the last paragraph of this section we shall discuss calculation of $\hat{g}$.) We also need an estimator $\hat{f}$ of the design density f.

To implement the bootstrap we first compute residuals $Y_i - \tilde{g}(X_i)$ and centre them, to obtain centred residuals $\hat{\varepsilon}_i$; then we resample $\varepsilon_1^*, \ldots, \varepsilon_n^*$ randomly with replacement from the centred residuals, and sample $X_1^*, \ldots, X_n^*$ randomly from the distribution with density $\hat{f}$; we bootstrap the model at (2.1) by taking $Y_i^* = \hat{g}(X_i^*) + \varepsilon_i^*$; and we compute the statistic T^*, being the version of T for the bootstrap data (X_i^*, Y_i^*) rather than the original data (X_i, Y_i). As the critical point for the test we use the α-level point, $\hat{t}(\alpha)$ say, of the distribution of T^* (conditional on the data). It is an approximation to the α-level point of the unconditional distribution of T. We reject H_0 at level α if $T > \hat{t}(\alpha)$. (We note that in the presence of heteroskedasticity of unknown form, the "wild" or "external" bootstrap can be used, see Wu, 1986.)

Under the null hypothesis the conditional distribution of $n^{1/2}\, T^*$ has the same weak limit, U say, as the unconditional distribution of $n^{1/2}\, T$, noted at (4.3). That is, if $g \in \mathscr{G}_0$ then

$$P_g(n^{1/2}\, T^* \leqslant x \,|\, \mathscr{D}) \to P(U \leqslant x), \tag{4.5}$$

uniformly in x, where $\mathscr{D}$ denotes the set of data $(X_1, Y_1), \ldots, (X_n, Y_n)$ generated by the model at (2.1). (See below for discussion of regularity conditions, and Delecroix et al., 2002 for a rigorous treatment in the single index case.) In particular, $n^{1/2}\, \hat{t}(\alpha)$ converges in probability to the α-level point of the distribution of U. Therefore, the bootstrap-calibrated test has asymptotically correct level. This remains true if we allow g to vary with n, converging to $g_0 \in \mathscr{G}_0$ as $n \to \infty$. This is the setting of local alternative hypotheses, to be discussed in Section 5. In the case of converging local alternatives the limiting distribution of $n^{1/2}\, T$, referred to above, is that where $g = g_0$.

Under a fixed alternative hypothesis the limit of the conditional distribution of $n^{1/2}\, T^*$ is the same as the limit that the unconditional distribution of $n^{1/2}\, T$ would have if the true g were g_1, the limit of $\hat{g}$. Therefore $\hat{t}(\alpha) = O_p(n^{-1/2})$. However, if the null hypothesis is not true, and if the alternative g is fixed, then $g \neq g_1$ and $n^{1/2}\, T$ diverges to infinity with probability 1. Therefore, the probability of rejecting H_0 converges to 1 in the case of a fixed alternative.

Finally, we discuss regularity conditions. We assume data are generated by model (2.1), where the variables $X_1, \ldots, X_n$ are independent and identically distributed with a density f which is bounded away from zero on the support of functions in $\mathscr{G}$. Regularity conditions needed on f depend on context. For instance, in the case of Example 1, if we use $\widehat{\psi}(\lambda|g)$ at (3.9) to estimate $\psi(\lambda|g)$ at (3.8) then we require only the conditions given below (3.9). If we have to use the density correction W_i,

however, then the smoothness assumed of f should increase with dimension, as noted below (3.7). In each of our examples no more than two continuous derivatives are required of g in order for the asymptotic properties discussed above to hold.

The estimators $\tilde{g}$, $\hat{g}$ and $\hat{f}$ should be constructed using sufficient smoothing to ensure convergence of these derivatives of the estimator to those of their limits. Let g_1 denote the in-probability limit of $\hat{g}$; it equals the true value of g if $g \in \mathscr{G}_0$, or (in the case of converging local alternatives) if $g \to g_0 \in \mathscr{G}_0$. The limiting distribution of T generally depends on the vector of first derivatives of g, through the term $\nabla\psi(\lambda, \theta_0)$ appearing at (4.4). Bounded second derivatives are important in controlling remainder terms when deriving the weak convergence theory in Section 4.1. Therefore, the smoothing parameters used to construct $\hat{g}$ should be chosen so that $\hat{g}$ and its first two derivatives converge uniformly, in probability, to their counterparts for g_1. The convergence rates are unimportant.

Assume too that $\tilde{g}$ converges uniformly to the true g; again the convergence rate is not important. Finally, we need the estimator $\hat{f}$ of f to be consistent and to have enough bounded derivatives to ensure condition (3.7) holds. As indicated there, we should assume uniform consistency of the first two derivatives, for dimensions up to three, and uniform consistency of the first four derivatives, for dimensions from four to seven.

Finally, we draw attention to methods for calculating $\hat{g}$ in the cases of our nine regression examples. In Examples 1 and 4, no estimator is needed, and in Example 2, simple linear regression suffices. In Example 3, spline methods (e.g. Wright and Wegman, 1980; Villalobos and Wahba, 1987; Ramsay, 1998; Kelly and Rice, 1990; Mammen and Thomas-Agnan, 1999), isotonic regression (e.g. Friedman and Tibshirani, 1984; Mammen, 1991) or kernel methods (e.g. Hall and Huang, 2001) can be used. For Example 5, a variety of estimators of additive regression models are available; see e.g. Stone (1985), Hastie and Tibshirani (1991), Linton and Nielsen (1995) and Linton (1997). For Example 7, estimators suggested by Robinson (1988) and Speckman (1988) are appropriate. Example 8 uses standard parametric methods. In Example 9, techniques proposed by Powell et al. (1989), Härdle et al. (1993), Ichimura (1993), Klein and Spady (1993) can be used.

5. Power against local alternatives

In developing theory for the power of our tests against local alternatives, we assume the null hypothesis is characterised as at (3.4) where $t(g)$ is defined by (3.2) and (3.3), and that the test statistic T is given by (3.5) or (3.6). We suppose for simplicity that the set Λ does not depend on θ. This was the case for all our examples, and instances where it does not hold may be treated using arguments similar to those below. Assume too that $\widehat{\psi}$ at (3.5) is a root-n consistent estimator of ψ, in the sense that for each $g_0 \in \mathscr{G}_0$ and any sequence $g = g_n$ converging to g_0,

$$\|\widehat{\psi}(\cdot, \cdot \mid g) - \psi(\cdot, \cdot \mid g)\| = \mathrm{O_p}(n^{-1/2}). \tag{5.1}$$

For the sake of definiteness, let us confine attention to the regression setting where data (X_i, Y_i) are generated by model (2.1). There, using arguments in Section 4.1, (5.1) follows quite generally from the assumption that the distribution of the pairs (X_i, ε_i) does not depend on g. In particular, when (4.1) holds or where instead (4.2) is valid, (5.1) follows via those results if the measure μ used to define the norm $\|\cdot\|$ is bounded.

Treat $\psi(\lambda, \theta|\cdot)$ as a functional from $\mathscr{G}$ to the real line, and suppose the Gâteaux derivative of the functional exists for perturbations away from g_0 in the direction of g_1:

$$\psi(\lambda, \theta|g_0, g_1) = \lim_{\delta \to 0} \delta^{-1} \{\psi(\lambda, \theta|g_0 + \delta g_1) - \psi(\lambda, \theta|g_0)\}.$$

Put $\theta_0 = \operatorname{argmin}_\theta \|\psi(\cdot, \cdot|g_0)\|_\theta$, assumed unique; let $\nabla\psi(\lambda, \theta|g)$ denote the vector of partial derivatives of $\psi(\lambda, \theta|g)$ with respect to θ; and consider the constraint:

$$\inf_\omega \|\psi(\cdot, \cdot|g_0, g_1) + \omega^{\mathrm{T}} \nabla\psi(\cdot, \cdot|g_0)\|_{\theta_0} > 0 \tag{5.2}$$

or, if the role of θ is degenerate:

$$\|\psi(\cdot|g_0, g_1)\| > 0. \tag{5.3}$$

Proposition 1 below gives conditions under which the test based on T is capable of distinguishing local alternative hypotheses that are distant $n^{-1/2}$ from $\mathscr{G}_0$. We shall need the following smoothness conditions:

> The derivatives of $\psi(\lambda, \theta|g)$ with respect to θ and g are well defined and continuous, in neighbourhoods of θ_0 and g_0, respectively, uniformly with respect to λ; and $\psi(\lambda, \theta|g_0, g_1)$ is continuous in θ lying within a neighbourhood of θ_0, again uniformly in λ. (5.4)

Proposition 1. *Assume* (5.1) and (5.4) hold. *If the test based on T has asymptotic level* $\alpha < 1$, *then* (5.2) [*or, if the role of* θ *is degenerate*, (5.3)] *implies that the test is capable of detecting departures of size* $n^{-1/2}$, *in the direction of* g_1, *from* $g_0 \in \mathscr{G}_0$. *That is, if* $g = g_0 + n^{-1/2} c\, g_1$ *where* $g_0 \in \mathscr{G}_0$ *and* $c > 0$, *and if* (5.2) [*or* (5.3)] *is valid, then*

$$\lim_{c \to \infty} \liminf_{n \to \infty} P_g(\text{test rejects } \mathrm{H}_0) = 1. \tag{5.5}$$

Conversely, if the approximation to ψ *by* $\widehat{\psi}$ *is not better than* $\mathrm{O}_p(n^{-1/2})$, *in the sense that* $T - t$ *is not of smaller order than* $n^{-1/2}$ (*uniformly in g in a neighbourhood of* $g_0 \in \mathscr{G}_0$), *then* (5.2) *is necessary for* (5.5).

Of course, we want the test to be able to distinguish local alternatives that are not in $\mathscr{G}_0$. Hence, in view of Proposition 1, we would like the following property to be satisfied:

> Condition (5.2) [or (5.3)] holds for all $g_0 \in \mathscr{G}_0$ and all $g_1 \in \mathscr{G}(g_0)$, (5.6)

where $\mathscr{G}(g_0)$ denotes the set of functions $g_1 \in \mathscr{G}$ that feature in the local alternatives $g_0 + \delta g_1$ that we consider. Conditions (5.6) and (3.4), and the result that ψ is

estimable root-n consistently in the sense of (5.1), are the main requirements we wish the class of functions $\psi(\cdot, \cdot \mid g)$ to fulfil.

Usually we take $\mathscr{G}(g_0)$ to be the set of all g_1 such that for all sufficiently small $|\delta|$, $g_0 + \delta g_1 \notin \mathscr{G}_0$. However, in some cases we narrow the field slightly; see the discussion following Proposition 2 below.

The validity of (5.6) is virtually guaranteed by (3.4), and so we really need only assume (3.4), as well as root-n consistency of $\widehat{\psi}$ for ψ. To appreciate why, first note that in the neighbourhood of $g_0 \in \mathscr{G}_0$ there generally exists a uniquely defined $\theta = \theta(g)$, say, at which $\|\psi(\cdot, \cdot|g)\|_\theta$ achieves its minimum. Assuming the functional ψ is smooth, and $g \to g_0$ in a smooth manner (let us take $g = g_0 + \delta g_1$, where $\delta \to 0$, for simplicity), the convergences of both $\theta(g)$ and $\|\psi(\cdot, \cdot|g)\|_{\theta(g)}$ to their respective limits, θ_0 and 0, are expressible in Taylor expansions:

$$\theta(g) = \theta_0 + \delta\,\theta_1 + \mathrm{o}(\delta), \quad \|\psi(\cdot, \cdot \mid g)\|_{\theta(g)} = \delta\,\beta_1 + \mathrm{o}(\delta),$$

where θ_1 and β_1 do not depend on δ. In general, neither θ_1 nor β_1 vanishes, and in particular, $t(g)$ converges to zero only linearly, i.e. at rate δ. Only in rare cases does $t(g)$ converge to zero at a faster rate than δ. Our next result shows that (5.6) is equivalent to $t(g)$ converging to zero at no faster than a linear rate.

Proposition 2. *Assume* $g_0 \in \mathscr{G}_0$ *and* $g_1 \in \mathscr{G}(g_0)$, *and* (5.4) *holds. Put* $g = g_0 + \delta g_1$. *Then* (5.6) *is equivalent to*: $|\delta| = \mathrm{O}\{t(g)\}$ *as* $\delta \to 0$.

Next, we show that (5.6) holds for each of Examples 1–10. In the context of Example 1, suppose $g_0 + \delta g_1 \notin \mathscr{G}_0$ for all sufficiently small δ, and that g_1 is continuous. Then g_1 cannot vanish identically. Now, $\psi(\lambda|g_0, g_1) = \int_\lambda g_1$, which implies (5.3) and hence (5.6). For Example 2, assume $g_0 + \delta g_1 \notin \mathscr{G}$ where g_0 is a linear function and g_1 is not, and that the measure μ places strictly positive mass on each of $\lambda = 2, 3, \ldots$ but zero mass on $\lambda = 0, 1$. Then $\psi(\lambda|g_0, g_1) = \int_\lambda g_1 \phi_\lambda$, for which (5.3), and hence (5.6), is satisfied.

Next we treat Example 3. Suppose g_0 is nondecreasing and that both g_0 and g_1 have continuous derivatives. If g_0 has a flat part on an interval $\mathscr{J}$ and is strictly increasing elsewhere, then, excepting pathological cases where g_1 is also flat on $\mathscr{J}$ but approaches the ends of $\mathscr{J}$ from an opposite direction to but with steeper gradient than g_0, $g_0 + \delta g_1$ is not nondecreasing if and only if g_1 is not nondecreasing on $\mathscr{J}$. In the latter case, if (in the notation of Example 3) $(\lambda_1, \lambda_3) \subseteq \mathscr{J}$, then

$$\psi(\lambda_1, \lambda_2, \lambda_3|g_0, g_1) = \min\left(\int_{\lambda_2}^{\lambda_3} g_1 - \int_{\lambda_1}^{\lambda_2} g_1\,,\, 0\right).$$

Since g_1 is not nondecreasing on $\mathscr{J}$ then this representation implies (5.3) and hence (5.6).

Treatment of Examples 4, 5 and 6 is analogous to that of Examples 3, 1 and 1, respectively. Examples 7–10 are similar to one another. In the case of Example 9, if $g_0(x) = \gamma(\theta_0 x)$ for all x, where γ is an r-variate function; and if g_1 cannot be written in this form for the same θ_0, modulo permutations of rows θ_0; then (5.2) holds, implying (5.6). In particular, (5.2) remains valid in the case of Example 9 even if both

g_0 and g_1 are r-variate index models, as long as the corresponding versions of θ cannot be expressed as row-wise permutations of one another.

6. Numerical properties

6.1. Simulations of tests of stochastic dominance

Let G_a and G_b be cumulative distribution functions with support contained in a finite interval, say $[0, \overline{x}]$. We say that G_a first-order stochastically dominates G_b if $G_a(x) \leqslant G_b(x)$ for all x; that G_a second-order stochastically dominates G_b if

$$\int_0^x G_a(t)\,\mathrm{d}t \leqslant \int_0^x G_b(t)\,\mathrm{d}t$$

for all x; and that G_a third-order stochastically dominates G_b if

$$\int_0^x \int_0^t G_a(s)\,\mathrm{d}s\,\mathrm{d}t \leqslant \int_0^x \int_0^t G_b(s)\,\mathrm{d}s\,\mathrm{d}t$$

for all x. Higher orders of stochastic dominance are defined similarly. Tests of these properties are of importance in the analysis of income distributions (hence the positive support) and in finance. Let $\Lambda = \{[0, x] \subseteq [0, \overline{x}]\}$. For first-order stochastic dominance, given $\lambda \in \Lambda$ define $\psi(\lambda) = \min[\{G_b(\lambda) - G_a(\lambda)\}, 0]$; for second-order, define

$$\psi(\lambda) = \min\left[\int_\lambda \{G_b(t) - G_a(t)\}\,\mathrm{d}t, 0\right]$$

and for third-order, put

$$\psi(\lambda) = \min\left[\int_\lambda \int_0^t \{G_b(s) - G_a(s)\}\,\mathrm{d}s\,\mathrm{d}t, 0\right].$$

Estimators $\widehat{\psi}(\lambda)$ are obtained by substituting empirical distribution functions, $\widehat{G}_a$ and $\widehat{G}_b$, for G_a and G_b, respectively.

Barrett and Donald (2003) considered Kolmogorov–Smirnov type tests of these properties. We compared the performance of our tests to theirs, as follows. Let x_a and x_b have distribution functions G_a and G_b, respectively. We considered the same five cases as Barrett and Donald (2003, pp. 85–86). Let z_a, z_b, $z_{b'}$ be independent N(0, 1) random variables with respective parameters (μ_a, σ_a), (μ_b, σ_b) and $(\mu_{b'}, \sigma_{b'})$. In each case, $x_a = \exp(\sigma_a z_a + \mu_a)$ with $\mu_a = 0.85, \sigma_a = 0.6$. For cases 1, 2 and 3, $x_b = \exp(\sigma_b z_b + \mu_b)$, with (μ_b, σ_b) equal to $(0.85, 0.6)$, $(0.6, 0.8)$ and $(1.2, 0.2)$, respectively. For cases 4 and 5,

$$x_b = I(u \geqslant 0.1)\exp(\sigma_b z_b + \mu_b) + I(u < 0.1)\exp(\sigma_{b'} z_{b'} + \mu_{b'}),$$

where the random variable u is uniformly distributed on $[0, 1]$. For these last two cases, $(\mu_b, \sigma_b, \mu_{b'}, \sigma_{b'})$ equal $(0.8, 0.5, 0.9, 0.9)$ and $(0.85, 0.4, 0.4, 0.9)$, respectively. In case 1, G_a is identical to G_b. In cases 2, 4 and 5, the null hypothesis that G_b

Table 1
Level and power of tests of stochastic dominance

$n_a = n_b$	50		500		50		500		50		500	
	First-order				Second-order				Third-order			
Case 1												
Level	H_0: true				H_0: true				H_0: true			
	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT
.01	.009	.019	.013	.013	.013	.017	.010	.012	.019	.019	.014	.013
.05	.035	.035	.049	.052	.057	.055	.053	.050	.053	.053	.052	.050
.10	.085	.108	.099	.099	.117	.108	.091	.094	.106	.103	.097	.097
Case 2												
Level	H_0: false				H_0: false				H_0: false			
	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT
.01	.288	.213	1.0	1.0	.116	.130	.882	.861	.147	.178	.884	.957
.05	.536	.447	1.0	1.0	.284	.305	.992	.971	.327	.383	.969	.989
.10	.661	.598	1.0	1.0	.434	.451	.998	.988	.476	.526	.987	.994
Case 3												
Level	H_0: false				H_0: true				H_0: true			
	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT
.01	.002	.009	.506	.857	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.05	.016	.053	.868	.991	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.10	.036	.149	.960	.999	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Case 4												
Level	H_0: false				H_0: false				H_0: false			
	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT
.01	.024	.041	.271	.390	.039	.040	.286	.293	.038	.037	.279	.271
.05	.093	.131	.479	.637	.127	.132	.535	.542	.127	.116	.506	.496
.10	.156	.238	.625	.753	.234	.236	.682	.683	.228	.227	.662	.647
Case 5												
Level	H_0: false				H_0: false				H_0: false			
	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT	KS	IT
.01	.037	.099	.790	.955	.068	.072	.776	.729	.065	.057	.634	.563
.05	.141	.257	.939	.997	.236	.222	.920	.906	.202	.185	.856	.791
.10	.238	.399	.982	1.00	.362	.335	.972	.953	.307	.289	.915	.889

The notation "KS" refers to the Kolmogorov–Smirnov tests in Barrett and Donald (2003), and "IT" refers to the integral test.

stochastically dominates G_a is false at all three orders. In case 3, G_b fails to stochastically dominate G_a only at first order (see Table 1).

We drew samples of size $n_a = n_b = 50$ or 500 from G_a and G_b. To apply the bootstrap, we took samples from the pooled dataset of $n_a + n_b$ observations, and simulated the distribution of test statistics under the null hypothesis. Table 1 summarises the results of our simulations. There, "KS" refers to the Kolmogorov–Smirnov type statistic proposed by Barrett and Donald (2003, p. 82) (see also Abadie, 2002); and "IT" refers to the integral test defined at (3.6). We performed 1000 simulations in each case and based critical values on 200 bootstrap samples. For tests of first-order stochastic dominance, the IT procedures displayed somewhat more power than the KS-type tests. For tests of second- and third-order stochastic dominance, the IT and KS procedures had similar power.

6.2. *Simulations of specification tests*

We compared the size and power properties of the integral test of Example 8 to their counterparts for residual regression tests. In the first set of simulations, we adopted the model used by Härdle and Mammen (1993), Li and Wang (1998) and Dette (1999). We found that the integral test generally outperformed the residual regression test.

In the second set of simulations we implemented the model in Horowitz and Spokoiny (2001) which incorporates a rapidly changing alternative. Upon the suggestion of a referee, we used least squares cross-validation to select the smoothing parameter for the residual regression test. In this case, we found the residual regression test to be more powerful than the integral test against the rapidly changing alternative.

6.3. *Empirical application: equivalence scale estimation and testing of equality of regression functions*

A classic problem in welfare and development economics is estimation of equivalence scales. To define the problem, suppose a couple needs, say, \$36,000 to achieve the same living standard as a single person earning \$20,000. Then the equivalence scale is said to be 1.8. One approach to estimating this quantity is to first estimate food Engel curves for various family types. Equivalence scales are estimated from the horizontal distances between these curves. (The basic intuition is that families achieve similar levels of well-being if they expend comparable shares of income on food.) In order that the procedure be valid, it is first necessary to test whether the curves are horizontal translations of one another.

Let X denote the logarithm of monthly income, and A and B be the numbers of adults and children, respectively. We used a South African survey dataset (see Yatchew et al., 2003), consisting of 4949 observations and comprised of 12 family types, represented by $A = 1, 2, 3$ and $B = 0, 1, 2, 3$. Fig. 1 illustrates estimated Engel curves for food for four common family types. Let G be the class of functions of the form $g(X, A, B)$. The null hypothesis H_0 consists of functions of the form

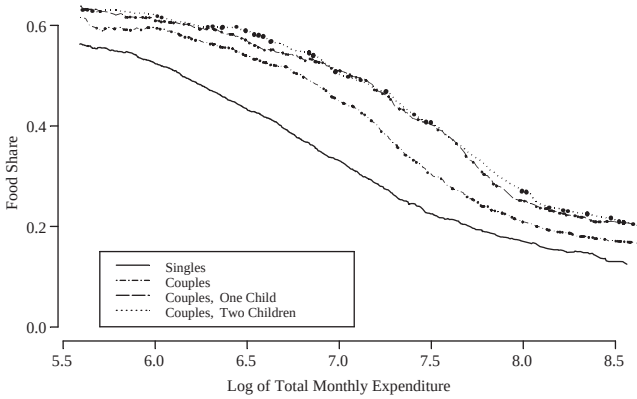

Fig. 1. Food Engel Curves.

$g_0\{X - \theta_1 \log(A + \theta_2 B)\}$, where θ_1, θ_2 lie between 0 and 1. The term $\Delta = (A + \theta_2 B)^{\theta_1}$ is interpreted as the equivalence scale, where θ_1 reflects the scale economies of living together and θ_2 measures the adult equivalence of a child. See Yatchew et al. (2003) for additional details.

Define $X^*_\theta = X - \theta_1 \log(A + \theta_2 B)$. The set Λ_θ is indexed by (a, b, A, B) where, for each vector (A, B), Λ_θ includes all intervals $[a, b]$ on which X^*_θ has positive support. Define

$$\psi(a, b, A, B, \theta \mid g) = \int_a^b g(X^*_\theta, A, B)\, \mathrm{d}X^*_\theta - \int_a^b g_0(X^*_\theta)\, \mathrm{d}X^*_\theta,$$

which we estimate using

$$\widehat{\psi}(a, b, A, B, \theta \mid g) = \sum_{\substack{i:\, a < X^*_{\theta i} < b \\ A_i = A,\, B_i = B}} 2\, Y_i\, D_1(X^*_{\theta i} \mid A_i, B_i) - \sum_{i\,:\, a < X^*_{\theta i} < b} 2\, Y_i\, D_1(X^*_{\theta i}),$$

where $D_1(X^*_{\theta i} | A_i, B_i)$ denotes the distance to the nearest value of $X^*_{\theta i}$ within the subset of the data that has A adults and B children, and $D_1(X^*_{\theta i})$ is the corresponding value for the full dataset. We obtained $(\hat{\theta}_1, \hat{\theta}_2) = (0.76, 0.80)$. The implied equivalence scale is 1.69 for couples, relative to singles, and 2.65 for couples with two children. The test statistic for the underlying hypothesis had the value 6.79, which is strongly significant; the 1% bootstrap critical value was 4.20. Therefore, one would reject the null hypothesis that there exist horizontal translations of the form $\theta_1 \log(A + \theta_2 B)$, which would cause the 12 individual regression curves to be superimposed. In contrast, when we restricted the data to singles and couples with no children (see Fig. 1), the test statistic was 0.01058, with a bootstrap probability estimate of 60%. The implied equivalence scale was in this case 1.65.

7. Theoretical arguments for Section 4

7.1. Proof of Proposition 1

Let $\hat{t} = \hat{t}(\alpha)$, computed from data, denote the critical point for the test. That is, the test amounts to rejecting H_0 if $T > \hat{t}$. Since the asymptotic level of the test is strictly less than 1 then, in view of (5.1), $\hat{t} = O_p(n^{-1/2})$ as $g \to g_0$. Property (5.1) also implies that

$$T(g) \geqslant t(g) - \|\widehat{\psi}(\cdot, \cdot \,|\, g) - \psi(\cdot, \cdot \,|\, g)\| \geqslant t(g) - O_p(n^{-1/2}),$$

uniformly in $|c| \leqslant n^{1/2}\,\delta$, for some $\delta > 0$. Therefore, (5.5) will follow if we show that

$$\lim_{c \to \infty} \liminf_{n \to \infty} n^{1/2}\, t(g) = \infty. \tag{7.1}$$

For (μ) almost all λ, and θ in a neighbourhood of θ_0,

$$\begin{aligned}\psi(\lambda, \theta|g) &= \psi(\lambda, \theta|g_0) + n^{-1/2}\, c\, \psi(\lambda, \theta|g_0, g_1) + o(n^{-1/2}) \\ &= n^{-1/2}\, c\, \psi(\lambda, \theta_0|g_0, g_1) + (\theta - \theta_0)^{\mathrm{T}}\, \nabla\psi(\lambda, \theta_0|g_0) \\ &\quad + o(n^{-1/2} + \|\theta - \theta_0\|).\end{aligned}$$

Therefore,

$$n^{1/2}\, t(g) = \inf_{\omega} \|c\, \psi(\cdot, \cdot \,|\, g_0, g_1) + \omega^{\mathrm{T}}\, \nabla\psi(\cdot, \cdot \,|\, g_0)\|_{\theta_0} + o(1). \tag{7.2}$$

Result (7.1) follows from this property and (5.2). Conversely, if the approximation to ψ by $\widehat{\psi}$ is not faster than $n^{-1/2}$ then for constants $C, \delta > 0$ the probability that t exceeds $Cn^{-1/2}$ exceeds δ for all sufficiently large n. Hence (5.5) implies (7.1), which, in view of (7.2), implies (5.2).

7.2. Proof of Proposition 2

The argument leading to (7.2) shows that in the context of Proposition 2,

$$t(g) = |\delta| \inf_{\omega} \|\psi(\cdot, \cdot \,|\, g_0, g_1) + \omega^{\mathrm{T}}\, \nabla\psi(\cdot, \cdot \,|\, g_0)\|_{\theta_0} + o(|\delta|)$$

as $\delta \to 0$. Proposition 2 follows immediately.

References

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

Anderson, N.H., Hall, P., Titterington, D.M., 1998. Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. Journal of Multivariate Analysis 50, 41–54.

Azzalini, A., Bowman, A., 1993. On the use of nonparametric regression for checking linear relationships. Journal of the Royal Statistical Society Series B 55, 549–557.

Baltagi, B.H., Hidalgo, J., Li, Q., 1996. A nonparametric test for poolability using panel data. Journal of Econometrics 75, 345–367.

Barrett, G.F., Donald, S.G., 2003. Consistent tests for stochastic dominance. Econometrica 71, 71–104.
Barry, D., 1993. Testing for additivity of a regression function. Annals of Statistics 21, 235–254.
Bierens, H., 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.
Bierens, H., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.
Bierens, H., Ploeberger, W., 1997. Asymptotic theory of integrated conditional moments. Econometrica 65, 1129–1151.
Bowman, A.W., Jones, M.C., Gijbels, I., 1998. Testing monotonicity of regression. Journal of Computational and Graphical Statistics 7, 489–500.
Delecroix, M., Hall, P., Vial, C., 2002. Tests for single index models that are powerful against local alternatives. Manuscript.
Dette, H., 1999. A consistent test for the functional form of a regression based on a difference of variance estimators. Annals of Statistics 27, 1012–1040.
Ellison, G., Ellison, S., 2000. A simple framework for nonparametric specification testing. Journal of Econometrics 96, 1–23.
Eubank, R., Hart, J., 1992. Testing goodness-of-fit in regression via order selection criteria. Annals of Statistics 20, 1412–1425.
Eubank, R., Spiegelman, C., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques. Journal of the American Statistical Association 85, 387–392.
Eubank, R., Hart, J., Simpson, D., Stefanski, L., 1995. Testing for additivity in nonparametric regression. Annals of Statistics 23, 1896–1920.
Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865–890.
Fan, J.Q., Lin, S.K., 1998. Test of significance when data are curves. Journal of the American Statistical Association 93, 1007–1021.
Friedman, J.H., Tibshirani, R.J., 1984. The monotone smoothing of scatterplots. Technometrics 26, 243–250.
Gijbels, I., Hall, P., Jones, M.C., Koch, I., 2000. Tests for monotonicity of a regression mean with guaranteed level. Biometrika 87, 663–673.
Gozalo, P., Linton, O., 2001. Testing additivity in generalized nonparametric regression models. Journal of Econometrics 104, 1–48.
Härdle, W., Mammen, E., 1993. Comparing nonparametric vs parametric regression fits. Annals of Statistics 21, 1926–1947.
Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single-index models. Annals of Statistics 21, 157–178.
Hall, P., Hart, J.D., 1990. Bootstrap test for difference between means in nonparametric regression. Journal of the American Statistical Association 85, 1039–1049.
Hall, P., Heckman, N.E., 2000. Testing for monotonicity of a regression mean by calibrating for linear functions. Annals of Statistics 28, 20–39.
Hall, P., Huang, L.-S. 2001. Nonparametric kernel regression subject to monotonicity constraints. Annals of Statistics, to appear.
Hall, P., Huber, C., Speckman, P.L., 1997. Covariate-matched one-sided tests for the difference between functional means. Journal of the American Statistical Association 92, 1074–1083.
Hart, J., 1997. Nonparametric Smoothing and Lack-of-Fit Tests. Springer, New York.
Hastie, T., Tibshirani, R., 1991. Generalized Additive Models. Chapman & Hall, London.
Hong, Y., White, H., 1995. Consistent specification testing via nonparametric series regression. Econometrica 63, 1133–1160.
Horowitz, J., Härdle, W., 1994. Testing a parametric model against a semiparametric alternative. Econometric Theory 10, 821–848.
Horowitz, J., Spokoiny, V., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69, 599–631.
Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics 58, 71–120.

Inglot, T., Kallenberg, W.C.M., Ledwina, T., 1994. Power approximations to and power comparison of smooth goodness-of-fit tests. Scandinavian Journal of Statistics 21, 131–145.

Inglot, T., Kallenberg, W.C.M., Ledwina, T., 2000. Vanishing shortcoming and asymptotic relative efficiency. Annals of Statistics 28, 215–238.

Ingster, Y.I., 1982. Minimax distinguishability of families of nonparametric hypotheses. Doklady Akademii Nauk SSSR (Russian) 267, 536–539.

Ingster, Y.I., 1993a. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. Mathematical Methods of Statistics 2, 85–114 Erratum Mathematical Methods of Statistics 2, 268.

Ingster, Y.I., 1993b. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. Mathematical Methods of Statistics 2, 171–189 Erratum Mathematical Methods of Statistics 2, 268.

Kelly, C., Rice, J., 1990. Monotone smoothing with application to dose–response curves and the assessment of synergism. Biometrics 46, 1071–1085.

Kendall, M., Stuart, A., 1979. The Advanced Theory of Statistics, vol. 2, fourth edn. Griffin, London.

King, E., Hart, J.D., Wehrly, T.E., 1991. Testing the equality of two regression curves using linear smoothers. Statistics and Probability Letters 12, 239–247.

Klein, R., Spady, R., 1993. An efficient semiparametric estimator for binary response models. Econometrica 61, 387–422.

Koul, H.L., Schick, A., 1997. Testing for the equality of two nonparametric regression curves. Journal of Statistical Planning and Inference 65, 293–314.

Kulasekera, K.B., Wang, J., 1997. Smoothing parameter selection for power optimality in testing of regression curves. Journal of the American Statistical Association 92, 500–511.

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. Econometric Theory 16, 576–601.

Lepski, O., Spokoiny, V., 1999. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. Bernoulli 5, 333–358.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. Journal of Econometrics 87, 145–165.

Linton, O.B., 1997. Efficient estimation of additive nonparametric regression models. Biometrika 84, 469–474.

Linton, O.B., Nielsen, J.P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82, 93–100.

Mammen, E., 1991. Estimating a smooth monotone regression function. Annals of Statistics 19, 724–740.

Mammen, E., Thomas-Agnan, C., 1999. Smoothing splines and shape restrictions. Scandinavian Journal of Statistics 26, 239–252.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press, Cambridge.

Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403–1430.

Ramsay, J.O., 1998. Estimating smooth monotone functions. Journal of the Royal Statistical Society Series B 60, 365–375.

Robinson, P.M., 1988. Root-*n* consistent semiparametric regression. Econometrica 56, 931–954.

Schlee, W., 1982. Nonparametric tests of the monotony and convexity of regression. In: Gnedenko, B.V., Puri, M.L., Vincze, I. (Eds.), Nonparametric Statistical Inference, vol. II. North-Holland, Amsterdam, pp. 823–836.

Speckman, P., 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society Series B 50, 413–446.

Spokoiny, V., 1996. Adaptive hypothesis testing using wavelets. Annals of Statistics 24, 2477–2498.

Stone, C., 1985. Additive regression and other nonparametric models. Annals of Statistics 13, 685–705.

Villalobos, M., Wahba, G., 1987. Inequality-constrained multivariate smoothing splines with application to the estimation of posterior probabilities. Journal of the American Statistical Association 82, 239–248.

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London.

Wooldridge, J., 1992. A test for functional form against nonparametric alternatives. Econometric Theory 8, 452–475.

Wright, I., Wegman, E., 1980. Isotonic, convex and related splines. Annals of Statistics 8, 1023–1035.

Wu, C., 1986. Jackknife bootstrap and other resampling methods in regression analysis. Annals of Statistics 14, 1261–1351.

Yatchew, A., 1992. Nonparametric regression model tests based on least squares. Econometric Theory 8, 435–451.

Yatchew, A., 1998. Nonparametric regression techniques in economics. Journal of Economic Literature XXXVI, 669–721.

Yatchew, A., 1999. An elementary nonparametric differencing test of equality of regression functions. Economics Letters 62, 271–278.

Yatchew, A., Bos, L., 1997. Nonparametric regression and testing in economic models. Journal of Quantitative Economics 13, 81–131.

Yatchew, A., Sun, Y., Deri, C., 2003. Efficient estimation of semiparametric equivalence scales with evidence from South Africa. Journal of Business and Economics Statistics 21, 247–257.

Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263–289.