Testing a Nonparametric Null Hypothesis Against a Nonparametric Alternative

Preliminary - comments are welcomed

Arie Beresteanu
Duke University

October 11, 2004

Abstract

This paper develops a test for a nonparametric null against a nonparametric alternative that takes into account maintained assumptions on the model. Economic theory may suggest that the functions in the model have certain properties but can be silent as to whether they have other properties of interest. In this case the researcher is interested in a procedure that takes the maintained assumptions of the model as given and can test whether it possesses additional qualities.

We define the null to be the model that imposes the maintained assumptions and a set of additional properties of interest on the functions. The alternative is a model where the maintained assumptions hold but the additional properties are violated. Both the null and the alternative are assumed to be nonparametric. Two sieve sequences are built to approximate the null and the alternative. These sieves are based on the shape restricted estimator developed in Beresteanu (2004).

JEL Classifications: C12 C14

Keywords: Testing, Shape restricted estimation, Sieve method.

1 Introduction

Hypotheses testing is one of the most important tasks for empirical work. Economic theory may suggest that the functions in the model have certain properties but can be silent as to whether they have other properties of interest. In this case, the researcher is interested in a procedure that takes the maintained assumptions of the model as given and can test whether it possesses additional qualities. For example, Beresteanu (2002) estimates a cost function imposing monotonicity with respect to outputs and concavity with respect to prices as the maintained assumptions on the cost function. Cost complementarities (submodularity of the cost function with respect to the vector of outputs) is the property being tested. In addition to being able to maintain basic assumptions on the model, it is important to have as less unnecessary assumptions as possible. Especially, functional form assumptions not dictated by economic theory are undesirable. This paper develops a test for a nonparametric null against a nonparametric alternative that takes into account the maintained assumptions of the model.

The tests discussed in the literature can be divided into three groups. The first group includes tests where both the null and the alternative hypotheses assume that the regression function belongs to a finite dimensional family of functions. The null hypothesis includes some restrictions on the parameters of the regression function. These testing problems are called here $P-P$ testing problems. The second category of tests includes test where the null hypothesis is parametric but the alternative is nonparametric, or in other words that the alternative is a infinite dimensional family of functions. These testing problems are called here $P-N$ testing problems. The idea behind the $P-N$ tests is that in $P-P$ we can reject the null against one specification of the alternative but accept it against another. In $P-N$ problems this is solved by having a nonparametric alternative. This paper discusses the third family of tests where both the null and the alternative are nonparametric. These tests are called $N-N$ tests. As mentioned above both the null and the alternative can be constrained with respect to the general family of functions on the support of the covariates but the null will include additional assumption.

The test proposed here is based on sieve estimators for both the null and the alternative. The null being the model which imposes the maintained assumptions and a set of additional properties of interest. The alternative is a model where the maintained assumptions hold but the additional properties are violated. Both the null and the alternative are assumed to be nonparametric. Two sieve sequences are built to approximate the null and the alternative. Although the method presented in this paper is very general, in this version I focus on a specific case - restricting the signs of the regression function's partial derivatives. The sieves constructed in this case are based on the shape restricted estimator developed in Beresteanu (2004). He employs a B-spline function basis where the coefficients assigned to the basis are constrained to assure that the sieve members satisfy the required properties.

One important feature of the shape restrictions discussed in this paper are their asymptomatic behavior. Beresteanu (2004) shows that shape restrictions do not affect the rate at which the estimator converges to the true regression function. As a result, both the estimator under the null and under the alternative has the same rate of convergence. In other words, the expected value of the distance between the estimator and the true regression function behaves as Cn^{-r}, where r is the rate of convergence. The difference between the two estimators will be in the constant C. Under the null, the estimator that takes into account the additional restrictions that are included in the null will be more efficient. This feature of the $N-N$ case is different from what is true in the $P-N$ case. In the last under the null it is possible to achieve $\sqrt{n}$ rate of convergence where under the alternative the rate is usually slower. The literature on $P-N$ testing includes three approaches. The first by Wooldridge (1992) considers building a sequence of non-nested tests. The second approach is described by Hong & White (1995) and uses nested testing. This is a simpler approach and allows the sieve sequence under the alternative to nest the sieve sequence under the null. The test statistic here is based on a comparison between the sum of squared errors under the null and under the alternative. The third kind of tests uses sample splitting and the null and the alternative are estimated each using only half of the sample (see Yatchew (1992)). This procedure is very costly especially in a nonparametric framework.

The literature on testing is vast. No attempt is being made to review all the literature on this topic here. The reader is referred to Zheng (1996) for a nice literature review. This paper differs from existing literature by assuming that both the null and the alternative are nonparametric. A specific estimator that can handle restrictions on the regression's partial derivatives is employed here but the method is more general and can include many other situations.

The paper proceeds in the following way. Section 2 outlines the basic assumptions and the testing problem discussed in this paper. Section 3 shortly describe the estimation of regression functions under shape restrictions. Section 4 described a test that allows the null to be nested in the alternative. The implications of shape restrictions to test power are discussed in Section 5. A short Monte-Carlo simulation is given in Section **??**. Section 6 concludes.

2 Statement of the Problem

This section outlines the assumptions made on the data generating process and on the estimation and testing procedures. The notations used in this paper are also established. We make the following assumptions.

Assumption 1: *The data generating process.* Let (Y, X, ε) be random variables in $\Re \times S \times \Re$ such that,
(i) $S \subset \Re^k$ is a compact set being the support of X.
(ii) $E(Y|X = x) = \theta(x)$ where θ is a finite Borel-measurable function on S and write the model as

$$Y = \theta(X) + \varepsilon.$$

(iii) Only Y and X are observed, their joint distribution is denoted by $\mathbb{P}$ and the marginal distribution of X is denoted by $\mathbb{P}_X$.
(iv) ε and X are statistically independent. The marginal distribution of ε is denoted by F_ε and ε has 4 finite moments.

The test statistics described below are based on independent and identically distributed observations $\{(x_i, y_i) : i = 1, 2, ...\}$ taken from the joint distribution $\mathbb{P}$. Next we state the assumptions on the regression model.

Assumption 2: *The functions space.* The regression function $\theta \in \Theta$ which is a family of measurable functions on S. The functions in Θ are square integrable with respect to $\mathbb{P}$, i.e. $\int_S [f(x)]^2 \, d\mathbb{P}(x) < \infty$ for all $f \in \Theta$.

The purpose of this paper is to develop a test for the following hypotheses

$$\begin{array}{c} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{array} \tag{1}$$

where $\Theta_0 \subset \Theta$ and $\Theta_1 = \Theta \backslash \Theta_0$ are two subclasses of measurable functions on S. When both Θ_0 and Θ are finite dimensional parametric families and the null hypothesis involves linear restrictions on the parameters of the functions in Θ, classical tests can be used. These tests include, among others, Wald tests, likelihood ratio tests and Lagrange multiplier tests for equality restrictions (see Amemiya (1985)). Analogous tests were developed by Gourieroux, Holly & Monfort (1982), Kodde & Palm (1986), Wolak (1989), Andrews (1998) and Meyer (2003) for linear inequality constraints. If Θ is nonparametric but Θ_0 is parametric then one of the various tests proposed for testing a parametric model against a nonparametric alternative can be used (e.g. Wooldridge (1992), Hardle & Mammen (1993), Hong & White (1995), Horowitz & Spokoiny (2001)). In this paper both Θ_0 and Θ_1 are nonparametric.

In the literature on testing a parametric null against a nonparametric alternative a sieve method is used to construct a sequence of tests each of which includes a parametric null against a parametric alternative. I use the same method to construct a series of tests when both the null and the alternative are nonparametric. Let $\{\Theta_{0,m}\}_{m=1..\infty}$ be a sieve sequence for Θ_0 and $\{\Theta_{1,m}\}_{m=1..\infty}$ be a sieve sequence for Θ_1. For each n we choose $m(n)$ such that $m(n) \to \infty$ as $n \to \infty$ and perform the following test

$$\begin{array}{c} H_{0,m(n)} : \theta \in \Theta_{0,m(n)} \\ H_{1,m(n)} : \theta \in \Theta_{1,m(n)}. \end{array} \tag{2}$$

Definition 1 *Define the* ***empirical measure*** $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ *where* δ_{X_i} *are the Dirac measures at the observation points. Given a set of functions* Θ *the empirical measure induces a map from* Θ *to* $\Re$ *by* $\mathbb{P}_n f = \int f d\mathbb{P}_n$. *For any* $f \in \Theta$ *define* $m_f = (Y - f(X))$ *and define the* ***empirical criteria function*** *to be the empirical process* $\mathbb{M}_n f = \mathbb{P}_n m_f = \frac{1}{n}\sum_{i=1}^{n} (Y_i - f(X_i))^2$ *and the limit process to be* $\mathbb{M} f = \mathbb{P} m_f = E(Y - f(X))^2$.

Using this definition we define the estimators and the best predictors in the following way

$$\begin{aligned} \hat{\theta}_{j,n} &= \arg \min_{f \in \Theta_{j,m(n)}} \mathbb{M}_n f, \ j = 0, 1 \\ \theta^*_{j,n} &= \arg \min_{f \in \Theta_{j,m(n)}} \mathbb{M} f, \ j = 0, 1 \\ \hat{\varepsilon}^i_{j,n} &= y_i - \hat{\theta}_{j,n}(x_i). \end{aligned}$$

To simplify the discussion we assume that the minimums above exist and yield a measurable functions. A generalization to cases where this is not true using outer limits and outer expectation is in Van der Vaart & Wellner (1996).

Let the scalar $r(\Im)$ denote the optimal rate of convergence that a nonparametric regression estimator can achieve if $E(Y|X)$ is known to be in $\Im$ in the sense of Stone (1980). It is clear that since $\Theta_0 \subset \Theta$, we have that $r(\Theta_0) \geq r(\Theta)$. In this version of the paper we focus on the case where $r(\Theta_0) = r(\Theta)$ (and thus also $r(\Theta_1) = r(\Theta)$). For simplicity we denote $r = r(\Theta_0) = r(\Theta)$.

3 Estimation under restrictions on partial derivatives

Although the framework described above is very general, this paper focuses on a certain type of shape restrictions - restrictions on partial derivatives of the regression function. These restrictions are very common in economics and among them one can find monotonicity, concavity and supermodularity. The sieve used here is based on a B-spline series estimator that allows imposing shape restrictions on the regression function in a convenient way. Other series estimators or sieves can be used. Therefore, some general conditions are stated and then the specifics of the B-spline estimator are reviewed.

Consider a sieve which is a sequence of finite dimensional linear spaces. These spaces include finite expansions using a base of functions such that the union of these spaces is dense in Θ_0 and in Θ. Assumption 3 bellow lists the basic requirements from the basis functions on which the sieve is based. In what follows we use m instead of $m(n)$ to simplify the notations. Places where m indicates a constant not depending on n are indicated specifically. Let $\Psi_m = \left\{\psi_{m,0}, ..., \psi_{m,m}\right\}$ be the $m+1$ functions that span both $\Theta_{0,m}$ and $\Theta_{1,m}$. The following assumption list the conditions on these bases.

Assumption 3: $\Theta_{j,m} = \left\{\tilde{\theta} : \tilde{\theta}(x) = \sum_{k=0}^{m} a_k \psi_{m,k}\right\} \cap \Theta_j$ for $j = 0, 1$. Let Ξ_n be the $n \times (m+1)$ matrix whose enteries are the basis functions $\left\{\psi_{m,0}, ..., \psi_{m,m}\right\}$ evaluated at the observation points $(x_1, ..., x_n)$. We assume that

(a) $\Xi_n' \Xi_n$ is nonsingular for all n sufficiently big $a.s$

(b) for each i, $\psi_{m,i} \Xi_n' \Xi_n \psi_{m,i} \to 0 \; a.s$

(c) $\exists \left\{\theta_{1,n}^* \in \Theta_{1,m}\right\}$ such that $\mathbb{P}\left(\theta_{1,n}^* - \theta\right)^2 \to 0$.

These assumptions are similar to the assumptions that guaranty asymptotic normality of series estimators as in Newey (1997).

Imposing shape restrictions on the regression function is discussed next. The sieve used here is the shape-restricted estimator sieve suggested in Beresteanu (2004). For each m, both $\Theta_{0,m}$ and $\Theta_{1,m}$ are based on a linear combination of a B-spline wavelet basis. For simplicity assume that X is a one dimensional covariate with support $[0,1]$. The basis functions are constructed by dividing the interval $[0,1]$ into m equally spaced grid points and having each element of the basis centered around one of the grid points. Beresteanu (2004) shows how to impose restrictions on an estimator which is based on these basis functions.

Let $\Psi_m^l = \left\{\psi_{m,0}^l, ..., \psi_{m,m}^l\right\}$ be the basis of B-spline functions of order l centered around the equidistant grid $\Gamma_m = (0, \frac{1}{m}, ..., 1)$. The set of splines defined by the basis Ψ_m^l is

$$S(\Psi_m^l) = \left\{ f : f(x) = \sum_{i=0}^{m} \beta_i \psi_{m,i}^l(x),\ |\theta_i| \leq \Delta \right\} \tag{3}$$

where Δ is some very big and known constant. Beresteanu (2004) shows that imposing restrictions on partial derivatives amounts to imposing linear inequality constraints on the coefficients $\{\beta_i\}_{i=0}^m$. If we denote by β the vector of coefficients then he shows how to build a matrix A^M such that if we impose $A^M\beta \leq 0$ the estimator is monotone another matrix which we denote by A^C can be used to impose convexity by imposing $A^C\beta \leq 0$. Close form formula is given for the restriction matrix for any restriction involving signing the partial derivative of the regression function. To see this we make the following definitions.

Definition 2 *A* ***differentiation matrix*** *of size p is a $p \times (p+1)$ matrix and is denoted by D_p and defined as*

$$D_p = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & -1 & 0 \\ 0 & \cdots & 0 & 0 & 1 & -1 \end{pmatrix}_{p\times(p+1)}$$

A function f is monotone increasing on the grid vector Γ_m if $f(0) \leq f(\frac{1}{m}) \leq ... \leq f(1)$ or in vector notations $D_m f(\Gamma_m)' \leq 0$ where $f(\Gamma_m)$ is interpreted as the vector of the values of the function f evaluated at the coordinates of the vector Γ_m. For concavity of the function f

on the grid vector Γ_m it needs to satisfy $D_{m-1}D_m f(\Gamma_m)' \leq 0$. This can be easily generalized to restrictions on any partial derivative and to a multi-dimensional setting as well using tensor product of B-splines.

The least squares estimator based on the function basis Ψ_m^l is:

$$\begin{aligned} & \min_{\beta} \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \left(y_i - \Psi_m^l(x_i)\beta' \right)^2 \\ s.t. & \\ & A_m^p \Psi_m(\Gamma_m)\beta' \geq 0 \end{aligned} \tag{4}$$

where A_m^p is the appropriate restriction matrix to impose non-negative p^{th} partial derivative on the expansion of B-spline function on the equidistant grid Γ_m. Theorem 1 in Beresteanu (2004) shows that for an appropriate choice of B-spline basis, $\hat{\beta}'\Psi_m(x)$ satisfies the shape restrictions over $[0, 1]$ and not only on the grid points. To avoid certain difficulties we assume the following on rank of A_m^p.

Assumption 4: The constraint in (4) is such that the $rank(A_m^p) \leq m + 1$.

4 Testing

Testing is based on the following property of the sieve sequence: $\int \left(\theta_{j,n}^*(x) - \theta(x)\right)^2 d\mu(x) \to 0$ as $n \to \infty$ for both $j = 0$ and $j = 1$ under the null. Under the alternative this is true only for $j = 1$ where as for $j = 0$ the distance between $\theta_{0,n}^*$ and θ converges to some positive number. An ideal test can include a situation where the alternative is shape restricted and the null contains additional restrictions which we want to test. For example, consider the case where we want to test whether the regression function is concave while maintaining the assumption that it is monotone. Under the null we solve $\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \Psi_m^l(x_i)\beta' \right)^2$ subject to $A_m^1\beta' \geq 0$ and $A_m^2\beta' \geq 0$. Under the alternative we minimize the same criteria function but only under the constraint that $A_m^1\beta' \geq 0$. In this version of the paper we consider only the case where the alternative is unconstrained and only the null hypothesis includes shape restrictions on the regression function.

4.1 Normally distributed errors

We start by looking at a simple problem and build from it. Consider testing $H_{0,m}$ against $H_{1,m}$ when m is a fixed number and the number of observations may tend to infinity. In addition we assume that the error term ε is identically distributed normal with know variance $\sigma^2 > 0$. Consider the following test statistic

$$F_n = \mathbb{M}_n \hat{\theta}_{0,n} - \mathbb{M}_n \hat{\theta}_{1,n}. \tag{5}$$

The assumption that the alternative is unconstrained means that only estimation under the null hypothesis involves linear inequality restrictions. For short notation let A_m be the matrix representing the linear inequality constraints and let $C_m = \{\beta | A_m \beta \geq 0\}$. Several algorithms have been suggested for finding the minimum of a least squares problem in a close convex polyhedral like C_m (see Fraser & Massam (1989) and Meyer (1999)).

Wolak (1989) develops the distribution of the test in (5).[1] To understand Wolak's result we first state the following result.

Theorem 1 *(Wolak (1989, Corollary 1)) Let $\hat{\mu} = \mu + v$ where μ is a P-dimensional vector and $v\tilde{\ }N(0, \Omega)$ where Ω is a $P \times P$ covariance matrix. Let $H_0 : \mu \geq 0$ and $H_1 : \mu \in \Re^P$ be a test. If $\tilde{\mu}$ is the solution for the quadratic programing*

$$\begin{aligned} &\min_{\beta} (\hat{\mu} - \beta)' \Omega^{-1} (\hat{\mu} - \beta) \\ &s.t \quad \beta \geq 0 \end{aligned} \tag{6}$$

then

a) $\mu = 0$ is the least favorable value of μ under the null hypothesis.

b) the test statistic $(\hat{\mu} - \tilde{\mu})' \Omega^{-1} (\hat{\mu} - \tilde{\mu})$ is distributed $\sum_{k=0}^{P} \chi_k^2 w(P, P-k, \Omega)$ where χ_0^2 is zero with probability 1 and $w(P, k, \Omega)$ is probability that $\tilde{\mu}$ has exactly k positive elements when $\mu = 0$.

The constraints cone C_m, can be divided into sub-cones each corresponding to a different set of constraints being binding and other nonbinding. The distribution of the test statistic above depends on the probabilities of each subset of constraints to be binding in a given sample. Next we use Theorem 1 to state the following result for constrained regressions.

[1] Robertson, Wright & Dykstra (1988) consideres $F_n = \left(\mathbb{M}_n \hat{\theta}_{0,n} - \mathbb{M}_n \hat{\theta}_{1,n}\right) / \mathbb{M}_n \hat{\theta}_{0,n}$. They show that the distribution of F_n is a mixture of Beta distributions. Let $B_{\alpha,\beta}$ be a random variable having a Beta distribution with parameters α and β. Then Robertson et al. (1988) show that $\Pr(F_n \leq a) = \sum_{d=0}^{k} \Pr(B_{\frac{k-d}{2}, \frac{d}{2}} \leq a) \Pr(D = d)$ where k is the number of independent lines in A and D is a random variable indicating the number of independent binding constraints and $B_{\alpha,0} \equiv 1$ and $B_{0,\beta} \equiv 0$.

Consider the quadratic programing problem in (4). Let $l = rank(A_m)$ and let B_m be such that $T_m = (A'_m \; B'_m)'$ is of full rank $m+1$. This is possible in light of Assumption 4. Define the following statistics

$$\begin{aligned} \bar{F} &= \min \left(Tb - T\hat{b}\right)' \left(T'^{-1}(X'\Omega^{-1}X)T^{-1}\right) \left(Tb - T\hat{b}\right) \\ & \quad s.t. \quad A_m b \geq 0 \end{aligned} \tag{7}$$

where $\hat{b}$ is the unconstrained minimizer. This quadratic programing problem can be rewritten as $\left(f - \hat{f}\right)' V^{-1}(f - \hat{f})$ where $f = Tb$, $\hat{f} = T\hat{b}$ and $V = \left(T'^{-1}(X'\Omega^{-1}X)T^{-1}\right)$.

Theorem 2 *(Wolak (1989, Theorem 2)) In the quadratic programing problem (4)*
a) any β^* such that $A_m\beta^{*\prime} = 0$ can serve as the least favorable for the constrained hypothesis
b)Let $W = \min_f \left(f - \hat{f}\right)' V^{-1}(f - \hat{f})$ such that $f_1 \geq 0$ where f_1 is the vector of the first $l$coordinates of f. Then W (and therefore also $\bar{F}$) is distributed $\sum_{k=0}^{l} \chi_k^2 w(l, l-k, \Sigma)$ where $\Sigma = A'_m(X'\Omega^{-1}X)A_m$.

The weights in Theorems (1) and (2) require knowing the weights $w(\cdot,\cdot,\cdot)$ for any number of restrictions and any covariance matrix. Several close form expressions are available in case the number of restrictions is not big. See for example the references in Wolak (1989, Theorem 2). Has he suggest these weights can be also computed using a Monte-Carlo method.

Let $\mu_m = \sum_{k=1}^{l} kw(l, l-k, \Sigma)$ be the expectation of the random variable $\sum_{k=0}^{l} \chi_k^2 w(l, l-k, \Sigma)$ and $2\mu_m$ be its variance. The test statistic we consider in this paper is

$$\tilde{F} = \frac{\bar{F} - \mu_m}{(2\mu_m)^{1/2}}.$$

Notice that this is a modification of the test statistic used in Hong & White (1995). The difference is that the expectation of $\bar{F}$ are no longer m and $2m$ respectively but μ_m and $2\mu_m$ since $\bar{F}$ is distributed as a mixture of chi-square distributions rather than a unique chi-square as is the case in Hong & White (1995). The advantage of this test statistic that it is symptomatically pivotal as the following theorem suggest.

Theorem 3 *Under assumptions $1-4$ above and for a fixed m, $\sqrt{n}\tilde{F} \xrightarrow{D} N(0,1)$.*

Further results including rates of convergence and the rate at which $m(n)$ should go to infinity are under development.

4.2 Non-normal distribution of errors - bootstrap test

This section deals with the case where ε_i's are not necessarily normally distributed. The course of action will be to show that a bootstrap procedure will work for this case. This development is useful also for the part where the errors are known to be normally distributed. The distribution of the test statistic in (7) is hard to compute and thus a bootstrap procedure is going to be useful in that case as well.

The following Monte-Carlo experiment was carried out. The model $Y = (\frac{1}{2} + 2X)^2 + \varepsilon$ with $\varepsilon \sim N(0,1)$ was employed. First, a null hypotheses which is correct was assumed and the Monte-Carlo experiment using 2000 repetitions computed the rejection probability which corresponds to the size of the test. The results are described in Table 1. Second a null hypotheses which is wrong was tested for the same samples and the rejection rate which corresponds to the power was computed. The results for the power of the test are described in table 2. In both cases the sieve mesh was 6 for $N = 400$ and 8 for $N = 800$.

Table 1: Rejection rates

Model and Hypotheses	$N = 400$	$N = 800$
DGP: $Y = (0.5 + 2X)^2 + \varepsilon$		
H_0: θ is convex and monotone	5.18	4.60
H_1: θ is non-convex and monotone		

Table 2: Rejection rates

Model and Hypotheses	$N = 400$	$N = 800$
DGP: $Y = (0.5 + 2X)^2 + \varepsilon$		
H_0: θ is concave and monotone	81.25	89.015
H_1: θ is non-concave and monotone		

5 Do shape restrictions increase power

In the introduction I was arguing that maintaining some assumptions on the model even on the alternative is interesting from economic perspective. Here I ask the question whether this has any statistical justification. To illustrate the question consider the following example. The researcher is interested in testing whether the regression function is concave. She is certain, however, that the regression function is monotone increasing. Two courses of actions are possible. In the first we define Θ_0 are the set of monotone and concave functions and Θ_1 as the set of the concave functions. The other option is to define Θ_0 as the set of concave functions and leave Θ_1 to be unrestricted. In other words, the monotonicity assumption is not being tested and the question is whether imposing it on the null and alternative assist in testing the assumption of interest.

6 Discussion and further research

The approach in this paper is very general. The null hypothesis is assumed to be nonparametric and represents the basic model plus some additional assumptions on the functions that need to be tested. Few interesting cases can be investigated. First, the null can be a space of functions of the same dimension as the alternative. For example the case discussed in the previous section. The set of functions which are both monotonic and convex is a subset of the set of monotonic functions but is of the same dimension. In this case it is plausible to assume that the sieves built for the null and the alternative will converge at the same rate to their respective spaces. Another interesting case is when both hypotheses belong to a nonparametric spaces but the null is of lower dimension. For example, the set of monotone and homogeneous of degree one functions (with more then one variable now) is included in the set of monotone functions but now the homogeneity assumption reduces the dimensionality of the null. In this case it would be reasonable to use different rates for the two sieve sequences. Another interesting question is whether including additional information improves the small sample power of the test. For example if we want to test whether the regression function is convex but we know that it is also monotone. We can set the test such that both the null and the alternative will satisfy monotonicity and test for convexity. We can also ignore monotonicity and just compare an estimator which is convex with an estimator which is unrestricted. Beresteanu (2004) shows that using prior information is valuable in small samples but not asymptotically. He also constructed a measure of information for these cases. It will be interesting to investigate whether these results follow to the testing question.

The small simulations done are very encouraging. I used the asymptotic properties of the test but it will be interesting to see whether a bootstrap test will preform better in small samples.

References

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

Andrews, D. W. (1998). Hypothesis testing with a restricted parameter space, *Journal of Econometrics* **84**: 155–199.

Beresteanu, A. (2002). Nonparametric analysis of cost complementarities in the telecommunications industry. Duke Economics Working paper 02-07.

Beresteanu, A. (2004). Nonparametric estimation of regression functions under restrictions on partial derivatives, Working Paper 06-04, Duke University.

Fraser, D. & Massam, H. (1989). A mixed primal-dual bases algorithm for regression under inequality constraints. application to convex regression, *Scandinavian Journal of Statistics* **16**: 65–74.

Gourieroux, C., Holly, A. & Monfort, A. (1982). Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters, *Econometrica* **50**(1): 63–80.

Hardle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *The Annals of Statistics* **21**(4): 1926–1947.

Hong, Y. & White, H. (1995). Consistent specification testing via nonparametric series regression, *Econometrica* **63**(5): 1133–1159.

Horowitz, J. L. & Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative, *Econometrica* **69**(3): 599–631.

Kodde, D. A. & Palm, F. C. (1986). Wald critiria for jointly testing equality and inequality restrictions, *Econometrica* **54**(5): 1243–1248.

Meyer, M. C. (1999). An extension of the mixed primal-dual bases algorithm to the case of more constraints than dimentions, *Journal of Statistical Planning and Inference* **81**: 13–31.

Meyer, M. C. (2003). A test for linear versus convex regression function using shape restricted regression, *Biometrika* **90**(1): 223–232.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics* **79**: 147–168.

Robertson, T., Wright, F. & Dykstra, R. (1988). *Order Restricted Statistical Inference*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons.

Stone, C. (1980). Optimal rate of convergence for nonparametric estimators, *The Annals of Statistics* **8**(6): 1348–1360.

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

Wolak, F. A. (1989). Testing inequality constraints in linear econometric models, *Journal of Econometrics* **41**(2): 205–235.

Wooldridge, J. M. (1992). A test for functional form against nonparametric alternatives, *Econometric Theory* **8**: 452–475.

Yatchew, A. (1992). Nonparametric regression tests based on least squares, *Econometric Theory* **8**(4): 435–451.

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