Uniform in Bandwidth Tests of Specification for Conditional Moment Restrictions Models

Pascal Lavergne        Pierre E. Nguimkeu
Toulouse School of Economics    Simon Fraser University
This version: August 20, 2010

Abstract

This paper addresses the issue of detecting misspecification on models defined by conditional moment restrictions (CMR). We propose new and practical specification tests built following either an approach à la Hausman (1978) or a score-type approach, see, e.g. Godfrey (1988). We show that both tests statistics are asymptotically equivalent, asymptotically chi-squared under the null hypothesis, and we derive their asymptotic power properties under a fixed and a sequence of local alternatives. The asymptotic behavior of the tests obtain uniformly in the bandwidth. To approximate the statistical behavior of our general tests in small and moderate samples, we propose bootstrap procedures to compute their critical values. To our knowledge this is the first general bootstrap proposed to date for specification testing in nonlinear models defined by CMR. The tests statistics and the bootstraps are simple to implement and simulation results show that they perform well in small samples.

Keywords: Conditional Moment Restrictions, Hypothesis Testing, Smoothing Methods, Empirical Processes, Generalized-Inverses, Bootstrap.

JEL Classification: C52, C12, C14, C15.

*Correspondence to: Pierre E.Nguimkeu, Dept of Economics, SFU, 8888 University Drive Burnaby BC V5A 1S6; Emails: npn1@sfu.ca; pascal.lavergne@univ-tlse1.fr
1 Introduction

This paper addresses the issue of detecting misspecification on models defined by conditional moment restrictions. Such models are often provided by the economic theory and are widely used in the econometric literature. A leading example is the theory of dynamic rational expectation models with time separable utility maximizing agents. In this theory, general equilibrium conditions typically predicts implications in terms of Conditional expectations. Other examples include models identified through instrumental variables, models defined by conditional mean and conditional variance without specific assumption on their distribution, nonlinear simultaneous equation models, and transformation models.

Estimation of those models have been intensively investigated in the literature. One of the most popular technique is the generalized method of moment (GMM) introduced by Hansen (1982), where a finite number of unconditional moment restrictions is derived from the conditional moment using the so-called instrumental variables (IV), which are arbitrary measurable functions of the conditioning variable. Subsequent techniques have been considered to provide more efficient and accurate estimators to Conditional Moment Restrictions Models. Chamberlain (1987) allowed for heteroskedasticity and showed that the semiparametric efficiency bound for CMR models can be attained. Robinson (1987), Newey (1990, 1993) discussed ways to obtain the semiparametric efficiency bound using nonparametric optimal instruments. Focusing on Smoothed Generalized Empirical Likelihood (GEL) methods, Donald, Imbens & Newey (2003), Kitamura, Tripathi & Ahn (2004), and Smith (2007a,b) provided one-step efficient estimators that does not require preliminary consistent estimators, whereas Antoine, Bonnal & Renault (2007) developed a three-step efficient estimator based on a smoothed euclidean Empirical Likelihood (EL).

The statistical behavior of the estimators derived from all the above methods rely on a user-chosen parameter, usually the smoothing or bandwidth parameter, which in turn depends on the sample data. In most cases, consistency and/or efficiency requires that this parameter be chosen such as to converge to zero when the sample size increases. Moreover, the practical choice of the parameter require further techniques that could be painful in empirical applications. Dominguez & Lobato (2006) introduced a class of estimators whose consistency does not depend on any user-chosen parameter, however, the semi-parametric efficiency bound cannot be attained with their procedure. In a recent work, Lavergne and Patilea (2008, henceforth denoted LP) proposed a new class of estimators obtained by Smooth Minimum Distance (SMD) estimation. Their theory provides a way to obtain $\sqrt{n}$-consistent and asymptotically normal estimators uniformly over a wide range of bandwidths including arbitrary fixed ones, that is, bandwidths that do not depend on the sample size. Moreover, for a vanishing bandwidth a semiparametrically efficient estimator for CMR can be obtained by their procedure. All the above mention estimation procedures rely on the crucial assumption that the Conditional Mo-
ment Restriction model under consideration is actually correctly specified. If the model is misspecified, the methods developed in the above theories may not hold and the resulting CMR estimators may be invalid. A central issue for the practitioner is therefore to check the validity of these moment restrictions upon which their estimation results crucially depend.

This paper proposes two alternative simple and practical procedures for testing the hypothesis that the model is correctly specified, that is, there exists a vector of parameter values that satisfies the conditional moments restrictions. One is a Hausman-type test based on the Mahalanobis distance between two LP estimators: a consistent and asymptotically efficient one- indexed by a vanishing bandwidth - and a consistent but inefficient one - indexed by a fixed bandwidth. The other is a score-type test based upon the gradient of the LP criterion computed with a vanishing bandwidth and evaluated at some consistent estimator derived with a fixed bandwidth. Both test statistics are asymptotically equivalent and asymptotically chi-squared distributed under the null. We also propose bootstrap methods to approximate these tests in small and moderate samples. The null distributions, validity of our bootstrap procedures, and local and global power properties of our tests are established uniformly in the bandwidth. Simulations show that the proposed specification tests have reasonable size and power performance in small and moderate samples.

Most of the existing specification tests that have been developed for CMR focused on testing a finite set of arbitrary unconditional moment restrictions implied by the conditional moment restrictions; see, e.g., the contributions of Newey (1985), Tauchen (1985) and Wooldridge (1990). However, these approaches have been criticized because transforming CMR into a finite set of unconditional moments potentially raises some identification issues. Dominguez and Lobato (2004), Delgado, Dominguez & Lavergne (2006) propose consistent specification tests based on a Cramer Von Mises criterion. Although their test enjoy optimality properties, its asymptotic distribution depends on the specific data generating process, thus making standard asymptotic inference procedures infeasible. Recent approaches like those of Tripathi & Kitamura (2003) and Otsu (2008) are based on empirical likelihood (EL) methods. EL methods however involve complex nonlinear optimization over many parameters making it difficult to implement in practice. In contrast, the tests we propose are simple to construct and easy to implement.

The rest of the paper is organized as follows. In Section 2, we present the framework and the proposed tests statistics. In Section 3, we discuss their asymptotic distribution and power properties. Bootstrap procedures to approximate the behavior of these tests are proposed in Section 4. Section 5 reports Monte Carlo simulations results showing that our tests possess satisfactory finite sample properties. Section 6 concludes whereas Section 7 gathers all the proofs and some of our technical formulas.
2 Framework and Tests Statistics

In this section, we describe the general framework for specification testing in CMR models. We also define our specification test statistics and give their rationale. We will follow the notations of Lavergne & Patilea (2008) as close as possible.

2.1 Basic framework

Define euclidean measurable space with the usual norm. For any matrix $A$, $||A|| = \sqrt{\text{tr}(AA')}$ denotes the Frobenius norm which reduces to the usual Euclidean norm when $A$ is a column vector. For a squared matrix $A$, $\lambda_{\text{min}}(A)$ and $\lambda_{\text{max}}(A)$ denote respectively the smallest and the biggest eigenvalue of $A$. For a real valued function $l(\cdot)$, $F[l(\cdot)]$ denotes its Fourier transform, $\nabla \theta l(\cdot)$ and $H_{\theta,\theta} l(\cdot)$ denote the $p$-column vector of first partial derivatives and $p \times p$ matrix of second derivatives of $l(\cdot)$ with respect to the $p$-dimensional vector $\theta$. If $l(\cdot)$ is a $r$-vector valued function, that is $l(\cdot) \in \mathbb{R}^r$, then $\nabla \theta l(\cdot)$ is rather the $p \times r$ matrix of first derivatives of the entries of $l(\cdot)$ with respect to the entries of $\theta$.

Suppose we have a random sample of independent observations $\{Z_i = (Y_i, X_i)\}_{i=1}^n$ of size $n$ on the $s$-and $q$-dimensional data vectors $X$ and $Y$; that is, $Y \in \mathbb{R}^s$, $X \in \mathbb{R}^q$, and $Z = (Y', X')' \in \mathbb{R}^{s+q}$, $s \geq 1$, $q \geq 1$. Typically, $X$ is continuously distributed with Lebesgue density function $f(\cdot)$ while $Y$ could be continuous, discrete or mixed.

Let $g(Z, \theta) = (g^{(1)}(Z, \theta), ..., g^{(r)}(Z, \theta))$ be a known $r$-vector of real valued borel-measurable functions of the observed data $Z$, and of the $p$-dimensional parameter vector $\theta$ which belongs to a $\mathbb{R}^p$-compact parameter set $\Theta \subset \mathbb{R}^p$, $p \geq 1$. The values of $Z$ are assumed to be related by an econometric model whose data generating process is given by the Conditional Moment Restriction

$$H_0 : \quad \mathbb{E}[g(Z, \theta_0)|X] = 0 \quad \text{almost surely (a.s.) for some} \quad \theta_0 \in \Theta. \quad (1)$$

Many econometric models are defined in this setup. In several contexts, the vector $g(Z, \theta)$ is interpreted as a residual vector from some nonlinear multivariate regression. In other contexts $\mathbb{E}[g(Z, \theta_0)|X]$ can be seen as first order partial derivatives of some stochastic dynamic optimization problem. To justify the use of a parametric model defined as in Equation (1), a specification test is needed. Thus, the null hypothesis $H_0$ to be tested is that the model defined by Equation (1) is correctly specified. In other words, the question can be formulated as follows: is there any $\theta_0 \in \Theta$ such that $\mathbb{E}[g(Z, \theta_0)|X] = 0$ a.s. ?

If one does not have any specific alternative model in mind, the alternative hypothesis to be tested would simply be that the above null hypothesis is false; that is:

$$H_1 : \quad 0 \leq \text{Pr}[\mathbb{E}[g(Z, \theta)|X] = 0] < 1 \quad \text{for all} \quad \theta \in \Theta \quad (2)$$
In other words, there exists no $\theta \in \Theta$ such that $E[g(Z, \theta) | X] = 0$ with probability 1.

We assume throughout the paper that the conditional moment restriction (eq1) uniquely identifies the parameter $\theta_0$. Our tests statistics use the Lavergne & Patilea (2008) smooth minimum distance (SMD) estimators for $\theta_0$ characterized by the CMR model (1). These estimators are obtained as solution of the minimization problem

$$
\min_{\theta \in \Theta} M_{n,h}(\theta, W_n) = \frac{1}{2n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} g'(Z_i, \theta)W_n^{-\frac{1}{2}}(X_i)W_n^{-\frac{1}{2}}(X_j)g(Z_j, \theta)K_{ij}
$$

with $K_{ij} = K\left(\frac{X_i - X_j}{h}\right)$, where $K(\cdot)$ is a multivariate kernel, and $h$ is the bandwidth parameter. $W_n(\cdot)$ is a sequence of $r \times r$ non-random positive definite weighting matrices with uniformly bounded spectral radius and converging pointwise to some positive definite matrix.

When the model is correctly specified, a $\sqrt{n}$-consistent and asymptotically normal estimator $\hat{\theta}_{n,d}$ can be obtained by solving Problem (3) for any fixed bandwidth $d$ and any positive definite weighting matrix $W_n(\cdot)$. The semiparametrically efficient SMD estimator $\hat{\theta}_{n,h}$ follows from a two-step procedure where the first step obtains a preliminary consistent SMD estimator for any bandwidth choice, and the second step uses a vanishing bandwidth $h$ and a kernel estimator $\hat{W}_n(\cdot)$ of $\text{Var}[g(Z, \theta_0) | X = \cdot]f(\cdot)$, the density-weighted conditional variance of $g(Z, \theta_0)$ as the weighting matrix. In particular, $\hat{W}_n(\cdot)$ can be obtained by

$$
\hat{W}_n(x) = W_n(x, \hat{\theta}_n) = \frac{1}{n} \sum_{k=1}^{n} g(Z_k, \hat{\theta}_n)g'(Z_k, \hat{\theta}_n)h^{-q}K((x - X_k)/h)).
$$

### 2.2 Hausman and Score types tests statistics

We construct two tests statistics, a Hausman and a Score type tests, subsequently denoted by $HW_{d,h}$ and $S_{d,h}$, to evaluate the hypothesis $H_0$. The rationale for these specification tests is explained as follows.

Following an approach a la Hausman (1978) we define the test statistics $HW_{d,h}$ as follows:

$$
HW_{d,h} = n(\hat{\theta}_{n,d} - \hat{\theta}_{n,h})\hat{Q}_{d}^{-1}(\hat{\theta}_{n,d} - \hat{\theta}_{n,h})
$$

where $\hat{Q}_{d}$ is a consistent estimator of $Q_d$, the asymptotic variance-covariance matrix of $\sqrt{n}(\hat{\theta}_{n,d} - \hat{\theta}_{n,h})$. Both $Q_d$ and $\hat{Q}_{d}$ are defined in Appendix A.
This test is based on a squared Mahalanobis distance between the semiparametrically efficient SMD estimator \( \hat{\theta}_{n,h} \) and the consistent but inefficient SMD estimator \( \tilde{\theta}_{n,d} \) which are jointly asymptotically normal. If the model is correctly specified, both estimators are consistent for \( \theta_0 \) so that their difference converge in probability to zero. Thus the proposed test has a simple chi-squared limiting distribution. In the presence of mis-specification, the distance between \( \hat{\theta}_{n,h} \) and \( \tilde{\theta}_{n,d} \) is expected to be nonzero even in large sample. Hence, significantly large values of \( HW_{d,h} \) are regarded as evidence that the null specification is not consistent with the data.

The idea of our Score-type test statistics is as follows. Let consider the gradient \( \nabla_{\theta}M_{n,h}(\cdot) \) of the SMD criterion indexed by a vanishing bandwidth \( h \) but evaluated at a consistent estimator \( \tilde{\theta}_{n,d} \) of \( \theta_0 \) indexed with a fixed bandwidth \( d \). If the model is correctly specified the quantity \( \nabla_{\theta}E_{M_{n,h}}(\theta_0) \) which is zero by first order optimization conditions of the population problem. Under the alternative, the population conditional moment implied by the model would be different from zero. Thus, the estimators \( \tilde{\theta}_{n,d} \) would converge to some pseudo true value \( \theta^*_d \) so that the above gradient evaluated at \( \theta^*_d \) would converge to a nonzero limit. This therefore suggests the construction of our Score-type test following a Godfrey (1988) approach as follows:

\[
S_{d,h} = n \nabla_{\theta}M_{n,h}(\tilde{\theta}_{n,d})\left[ H_{\theta\theta}M_{n,h}(\tilde{\theta}_{n,d})\tilde{Q}_dH_{\theta\theta}M_{n,h}(\tilde{\theta}_{n,d}) \right]^{-1}\nabla_{\theta}M_{n,h}(\tilde{\theta}_{n,d}),
\]

(6)

The regularity conditions surrounding the construction of \( HW_{d,h} \) and \( S_{d,h} \) given by formulas (5) and (6) are given in the next section. If the variance-covariance matrix \( Q_d \) is singular, then \( HW_{d,h} \) and \( S_{d,h} \) must be computed using a generalized-inverse \( \hat{Q}_d \) of \( Q_d \) as described in the next section.

3 Asymptotic Properties of the Tests

This section establishes the asymptotic behavior of \( HW_{d,h} \) and \( S_{d,h} \) under the null and alternative hypotheses. We also investigate the properties of the tests under sequences of local alternatives. We start by giving basic assumptions.

3.1 Assumptions

We now provide some regularity conditions under which the asymptotic theory of our specification tests statistics are derived.

Assumption 1. \( \theta_0 \) is the unique value in \( \text{int}(\Theta) \) satisfying Equation (1), that is, \( \text{E}[g(Z,\theta)|X] = 0 \text{ a.s implies } \theta = \theta_0. \)

Assumption 2. (i) The kernel \( K(\cdot) \) is a symmetric, bounded real-valued function, which integrates to one on \( \mathbb{R}^q, \int K(u)du = 1. \)
(ii) The class of all functions \((x_1, x_2) \mapsto K(\frac{x_1 - x_2}{h})\), \(x_1, x_2 \in \mathbb{R}^q\), \(h > 0\), is Euclidean for a constant envelope \(^1\).

(iii) The Fourier transform \(\mathcal{F}[K](\cdot)\) of the kernel \(K(\cdot)\) is strictly positive and Holder continuous with exponent \(a > 0\).

**Assumption 3.** \(W_n(\cdot)\) is a sequence of \(r \times r\) positive definite non-random weighting matrices with \(0 < \inf_{n,u} \lambda_{\min}(W_n(u)) \leq \sup_{n,u} \lambda_{\max}(W_n(u)) < \infty\), and have a symmetric positive definite pointwise limit \(W(\cdot)\), that is, \(W_n(u) - W(u) = o(1)\), for all \(u\).

Assumption 1 above ensures that the model under consideration is identified. The fact that \(\theta_0\) should belong to \(\text{int}(\Theta)\) will be needed when deriving asymptotic behavior of our estimators as one often need to make some first or second order Taylor expansions around the parameter value of interest in the interior of the parameter set. This also allows \(\theta_0\) to be an interior solution of the population analogue of the minimization problem. Assumption 2 (i) states conditions that the Kernel functions need to satisfy. As for Assumption 2 (ii), we refer to Nolan & Pollard (1987), Pakes & Pollard (1989) and Sherman (1994a) for the definition and properties of Euclidean families. The strict positivity of the Fourier transform of the kernel \(K(\cdot)\) mentioned in Assumption 2 (iii) is useful to establish the consistency of the SMD estimators under the null (see Lavergne & Patilea 2008). This condition is fulfilled for instance by products of the triangular, normal, Laplace or Cauchy densities, but also by more general kernels, including higher-order kernels taking possibly negative values. Assumption 3 insures that the weighting matrix \(W_n^{-1/2}(\cdot)\) is well defined and that its spectral radius is uniformly bounded.

Define the following functions: \(g_n(z, \theta) = W_n^{-1/2}(x)g(z, \theta)\), \(\tau(x, \theta) = \mathbb{E}[g(Z, \theta)|X = x]\), and \(\tau_n(x, \theta) = W_n^{-1/2}(x)\tau(x, \theta) = \mathbb{E}[g_n(Z, \theta)|X = x]\). The Assumptions stated below are regularity conditions on the function \(g(\cdot, \cdot)\), its first and second derivatives and their Fourier transforms.

**Assumption 4.** (i) The function \(\theta \mapsto \tau(x, \theta)\) is continuous for any \(x\), and the function \(x \mapsto \sup_{\theta} \| \tau(x, \theta) \| \ f(x)\) belongs to \(L^2 \cap L^1\)

(ii) The families \(\mathcal{G}_k = \{g^{(k)}(\cdot, \theta) : \theta \in \Theta\}\), \(1 \leq k \leq r\) are Euclidean for a squared integrable envelope \(G\) with \(\mathbb{E}G^2 < \infty\).

(iii) There exists a constant \(c > 0\) such that \(\mathbb{E} \| g(Z, \theta_1) - g(Z, \theta_2) \|^2 \leq c \| \theta_1 - \theta_2 \|\), for all \(\theta_1, \theta_2 \in \text{int}(\Theta)\).

(iv) The components of \(\nabla_\theta \tau(\cdot, \theta_1)f(\cdot)\) and of \(\mathbb{E}[g(Z, \theta_1)g'(Z, \theta_2)|X = \cdot]f(\cdot)\) are continuous in \(\theta_1, \theta_2 \in \text{int}(\Theta)\) and uniformly bounded in \(L^2 \cap L^1\).

\(^1\)We use the definition and properties of Euclidean families as described in Nolan & Pollard (1987), Pakes & Pollard (1989), and Sherman (1994a)
Assumption 4 (i) guarantees that $E_{M_{n,h}}(\theta, W_n)$ is a continuous function for both $\theta$ and $h$. Assumption 4 (ii) ensures that any family of functions obtained as products or linear combinations of $g(z, \theta)$ and its components is uniformly Euclidean for a squared integrable envelope.

Assumption 5. (i) All second partial derivatives of $g(z, \cdot)$ exist on $\text{int}(\Theta)$.
(ii) There exists a real valued function $H(\cdot)$ with $\mathbb{E} H^4 < \infty$ and some constant $a \in (0, 1]$ such that:
$$\| H_{\theta, \theta} g^{(k)}(Z, \theta_1) - H_{\theta, \theta} g^{(k)}(Z, \theta_2) \| \leq H(Z) \| \theta_1 - \theta_2 \|^a, \forall \theta_1, \theta_2 \in \text{int}(\Theta), k = 1, \ldots, r.$$
(iii) The non-random matrices $V_{n,h}(\theta) = H_{\theta, \theta} E_{M_{n,h}}(\theta)$ and $V_{n,0}(\theta_0) = \lim_{h \downarrow 0} V_{n,h}$ are defined such that $\inf_{n,h} \lambda_{\min}(V_{n,h}(\theta)) > 0$, $\forall \theta \in \text{int}(\Theta)$;

Assumption 5(iii) guarantees the non-singularity of the matrices $V_{n,h}(\theta)$ and therefore that of $H_{\theta, \theta} M_{n,h}(\theta)$.

Assumption 6. The components of $\nabla_{\theta} \tau_{n}(\cdot, \cdot) f(\cdot)$ and of $H_{\theta, \theta} \tau_{n}^{(k)}(\cdot, \cdot) f(\cdot)$, $1 \leq k \leq r$, belong to the Sobolev Space of functions $H^s$, endowed with the norm:
$$\| \phi \|^2_{H^s} = \int_{\mathbb{R}^d} (1 + \| t \|^2)^s | \mathcal{F}[\phi](t) |^2 dt,$$ with $s > \frac{3}{2}$. 

It should be emphasized that for the construction of our Hausman-Wald test statistic $HW_{d,h}$, Assumptions 5 and 6 - regarding the twice differentiability of $g(z, \cdot)$ - are not needed. The specification testing procedure therefore applies to a wider variety of models. As for the score-type test statistic $S_{d,h}$, the asymptotically efficient estimator $\hat{\theta}_{n,h}$ of $\theta_0$ needs not be computed. In fact, only $\tilde{\theta}_{n,d}$ - the preliminary estimator - and the first and second order derivatives of $g(z, \cdot)$ are used for this purpose. The test therefore has a considerable computational advantage as it does not require a second-step optimization of the objective function $M_{n,h}$.

3.2 Asymptotic null distribution

Let: $\mathcal{H}_n = \{ h_0 \geq h > 0 : nh^{4q/\alpha} \geq C \}$ and $\mathcal{H}'_n = \{ 1/\ln(n+1) \geq h > 0 : nh^{4q/\alpha} \geq C \}$ where $h_0 > 0$, $C > 0$, and $\alpha \in (0, 1)$ are arbitrary constants that may take different values in different contexts.

The asymptotic equivalence between the Hausman and the Score-type specification test statistics is given in the following Lemma.

---

2See Malliavin(1995,Section III.3) for more discussion. This assumption can be replaced by others weaker conditions as described in Lavergne and Patilea (2009, Appendix A, Lemma 7.5)
Lemma 1. Let Assumptions 1-6 hold. Then under $H_0$ and uniformly over $h \in \mathcal{H}_n$, $HW_{d,h} = S_{d,h} + o_p(1)$, for any fixed $d$.

This result tells us that when the model is correctly specified, both test statistics have the same asymptotic behavior. As we will see in the next sections, this is also true under alternatives. The following result is immediate.

Theorem 1. Let Assumptions 1-6 and $H_0$ hold. Then under $H_0$ and uniformly over $h \in \mathcal{H}_n$, the test statistics $HW_{d,h}$ and $S_{d,h}$ both converge in distribution to $\chi^2(p)$, for any fixed $d$.

The degree of freedom $p$ is the dimension of the parameter vector $\theta$ which is also the rank of the variance-covariance matrix $Q_d$. In the case where the asymptotic covariance matrix $Q_d$ is singular with rank $s < p$ the test statistics given by Equations (5) and (6) are still valid provided one replaces the inverse $Q_d^{-1}$ in the expressions of the tests by a modified inverse as proposed by Lutkepohl and Burda (1997) or a regularized inverse as proposed by Dufour & Valery (2009). Unlike using generalized-inverses as suggested in the earlier literature (see, e.g., Hausman & Taylor 1981, Holly 1982), these approaches are designed to build inverses that remain continuous so that the distributional behavior of the tests statistics built from them are not misleading. In our case the limiting null distribution of the tests $HW_{d,h}$ and $S_{d,h}$ so modified remain chi-squared but with $s$, $s < p$ degrees of freedom.

3.3 Asymptotic power properties against global alternatives

In this subsection, we analyze the power properties of our test; that is, its behavior under the alternative $H_1$, or its ability to reject the null hypothesis $H_0$ when it is false.

In the presence of misspecification, the population conditional moment $E[g(Z, \theta)|X]$ is different from zero for any value of the parameter $\theta$. In this case, SMD estimators $\hat{\theta}_{n,d}$ and $\hat{\theta}_{n,h}$ would typically converge to two different limits. Denote $\theta_d^*$ and $\theta_0^*$ the probability limits of $\hat{\theta}_{n,d}$ and $\hat{\theta}_{n,h}$ when the model is misspecified, and by $Q_d^*$ their asymptotic covariance matrix under the global alternative $H_1$.

Theorem 2. Suppose Assumptions 1-6 and $H_1$ hold, and $\theta_d^* - \theta_0^* \neq 0$. Then uniformly over $h \in \mathcal{H}_n$ and for any arbitrary constant $C$:

$$P_r[HW_{d,h} > C] \rightarrow 1 \text{ and } P_r[S_{d,h} > C] \rightarrow 1, \text{ as } n \rightarrow \infty$$

The condition $\theta_d^* - \theta_0^* \neq 0$ in Theorem 2 is crucial for the result to hold. In fact, if under the alternative $H_1$ the estimators $\tilde{\theta}_{n,d}$ and $\tilde{\theta}_{n,h}$ are such that $plim[\tilde{\theta}_{n,d} - \tilde{\theta}_{n,h}] = 0$, that
is, $\theta^*_n - \theta^*_0 = 0$, then $\text{plim} \nabla \theta M_{n,h}(\tilde{\theta}_{n,d}) = 0$ (see proofs section). In this case not much information about the specification of the model have been provided by the test. It is therefore important to examine how likely such situation would arise in practice. The power of the tests depends on how far the probability limit of $\tilde{\theta}_{n,d}$ is from the probability limit of $\hat{\theta}_{n,h}$ (or how far the probability limit of $\nabla \theta M_{n,h}(\tilde{\theta}_{n,d})$ is from zero). Luckily, given the flexibility of the uniform-in-bandwidth framework under which those estimators are derived, it is always possible to choose the fixed bandwidth $d$ such that the difference of the estimators $(\tilde{\theta}_{n,d} - \hat{\theta}_{n,h})$ (resp., the gradient $\nabla \theta M_{n,h}(\tilde{\theta}_{n,d})$) is expected to be large (resp., far from zero) if a certain type of misspecification in suspected. In our simulations, we explore various classes of alternatives, and find out that such situations where the condition is violated would hardly arise in practice.

### 3.4 Asymptotic power against local alternatives

We now consider a sequence of local alternatives that approaches the null hypothesis as $n$ increases.

$$H_{1,n}: \quad \mathbb{E}[g(Z, \theta_n)|X] = \frac{\delta(X)}{\sqrt{n}} \quad \text{almost surely (a.s.),}$$

where $\theta_n$ is a nonstochastic sequence of parameter values in $\Theta$ that converges to $\theta_0$, and $\delta: \mathbb{R}^d \to \mathbb{R}^r$ is a continuous and squared integrable function. The latter property is needed to derive the asymptotic behavior of the tests. Note that Condition (7) and the convergence of $\theta_n$ to $\theta_0$ also imply that the SMD estimators $\tilde{\theta}_{n,d}$ and $\hat{\theta}_{n,h}$ are $\sqrt{n}$-consistent for $\theta_n$. Thus, one can replace $\tilde{\theta}_{n,d}$ and $\hat{\theta}_{n,h}$ by $\theta_n$ in the expressions of $HW_{d,h}$ and $S_{d,h}$ without changing their asymptotic distribution under $H_{1,n}$.

Denote:

$$\eta_d = \mathbb{E} \left[ \nabla \theta \tau(X_1, \theta_0) W_{1/2}(X_1) W_{1/2}(X_2) \delta(X_2) d^{-q} K((X_1 - X_2)/d) \right],$$

$$\eta_0 = \mathbb{E} \left[ \nabla \theta \tau(X, \theta_0) \text{Var}^{-1} g(Z, \theta_0)|X] \delta(X) \right],$$

and

$$\mu_d = V_d^{-1} \eta_0 - V_0^{-1} \eta_0.$$

where $V_d$ and $V_0$ are the asymptotic covariance matrices both defined in Appendix A. We have the following result.

**Theorem 3.** Let Assumptions 2-6 hold. Then under the sequence of local alternatives $H_{1,n}$, the test statistics $HW_{d,h}$ and $S_{d,h}$ converge to a noncentral chi-square distribution $HW_{d,h} \overset{d}{\rightarrow} \chi^2(\mu_d^T Q_d \mu_d, p)$ and $S_{d,h} \overset{d}{\rightarrow} \chi^2(\mu_d^T Q_d \mu_d, p)$, uniformly over $h \in \mathcal{H}_n'$ for any fixed bandwidth $d$.

The test statistics $HW_{d,h}$ and $S_{d,h}$ therefore have nontrivial power against the sequence of local alternatives $H_{1,n}$.
All our results obtained above then show that a Hausman-type specification test based on the difference $\hat{\theta}_{n,d} - \hat{\theta}_{n,h}$ would be asymptotically equivalent to a score-type test based on the gradient $\nabla_{\theta} M_{n,h}(\hat{\theta}_{n,d})$.

4 **Bootstrap Test**

Since asymptotics sometimes fail to accurately reflect the behavior of test statistics in small or moderate samples, bootstrap is an alternative way to better approximate their distribution (see Hardle & Mammen (1993), Li & Wang (1998) and the references therein). In what follows, we describe the bootstrap method that we propose to compute critical values for CMR specification tests and we formally establish their validity.³

A popular procedure that have been shown to produce accurate results for specification test statistics is the wild bootstrap method; see e.g., Mammen (1992), Liu (1988), and Jin, Ying & Wei (2001). Instead of randomly sampling with replacement from the pooled data $\{Z_i, i = 1 \ldots n\}$ this method suggests to generate an artificial sample $\{g_b(Z_i, \theta) = \omega_i g(Z_i, \theta), i = 1 \ldots, n\}$ from the observations $\{g(Z_i, \theta), i = 1, \ldots, n\}$ that allows us to define the perturbed criterion:

$$M_{n,h}^{b}(\theta, W_n) = \frac{1}{2n(n-1)b^2} \sum_{1 \leq i \neq j \leq n} \omega_i \omega_j g'(Z_i, \theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta) K\left(\frac{X_i - X_j}{h}\right)$$

(11)

Note that the wild bootstrap weights $\{\omega_i, i = 1 \ldots n\}$ should be $n$ independent identical copies of a known positive random variable $\omega$ that satisfies the condition $E(\omega) = Var(\omega) = 1$.

Since $\omega$ is independent of the original data, it is easy to see that the above condition on $\omega$ implies $E[g_b(Z, \theta)|X] = E[g(Z, \theta)|X]$ and $Var(g_b(Z, \theta)|X] = Var[g(Z, \theta)|X]$ so that the basic properties of the original criterion are preserved.

To further improve the rate of convergence of the bootstrap estimate, one could further impose the supplementary condition that $\text{Skewness}(\omega) = 1$ as introduced by Liu (1988), and Hardle and Mammen (1990). With the new criterion defined by Equation (11), we repeat the optimization process by estimating $\hat{\theta}_{n,d}^b$, the bootstrap SMD estimator with fixed bandwidth $d$ and $\hat{\theta}_{n,h}^b$, the efficient one with vanishing bandwidth $h \in H_n': We can then compute the bootstrap version of our Hausman-type test by

$$HW_{d,h}^b = n(\hat{\theta}_{n,d}^b - \hat{\theta}_{n,d} + \hat{\theta}_{n,h} - \hat{\theta}_{n,h})/\hat{Q}_d^{-1}(\hat{\theta}_{n,d} - \hat{\theta}_{n,d} + \hat{\theta}_{n,h} - \hat{\theta}_{n,h})$$

(12)

³The fact that our test statistic is asymptotically pivotal, that is, it has a conventional asymptotic distribution, gives us the right to actually apply the bootstrap (for a discussion on situations where bootstrap can or cannot be apply, see, e.g., Horowitz (2001))
where, \( Q^b \) is the bootstrap counterpart of \( Q \), and \( \tilde{\theta}_{n,d} \) and \( \tilde{\theta}_{n,h} \) are non-bootstrap SMD estimators. The process is repeated a large number of times, say \( B \) times, to obtain an empirical distribution of the \( B \) bootstrap test statistics \( \{HW^b_{d,h,j}\}_{j=1}^B \). This bootstrap empirical distribution is then used to approximate the distribution of the test statistic \( HW_{d,h} \) under the null, allowing to calculate the critical values empirically. Typically, one rejects \( H_0 \) at \( \alpha \) level if \( HW_{d,h} > C^{H}_{\alpha B} \), where \( C^{H}_{\alpha B} \) is the upper \( \alpha \)-percentile of the empirical distribution \( \{HW^b_{d,h,j}\}_{j=1}^B \).

We can also compute the bootstrap version of our score-type test statistics using the following formula:

\[
S^b_{d,h} = n\left[ \nabla_{\theta}^\prime M^b_{n,h}(\tilde{\theta}^b_{n,d}) - \nabla_{\theta} M^b_{n,h}(\tilde{\theta}_{n,d}) \right] \left[ H_{\theta \theta} M^b_{n,h}(\tilde{\theta}^b_{n,d}) \tilde{Q}^b_{d,h} H_{\theta \theta} M^b_{n,h}(\tilde{\theta}^b_{n,d}) \right]^{-1} \times \\
\times \left[ \nabla_{\theta} M^b_{n,h}(\tilde{\theta}^b_{n,d}) - \nabla_{\theta} M^b_{n,h}(\tilde{\theta}_{n,d}) \right] 
\]

(13)

Likewise, an empirical distribution of \( B \) bootstrap test statistics \( \{S^b_{d,h,j}\}_{j=1}^B \) can be obtained. One rejects \( H_0 \) at \( \alpha \) level if \( S^b_{n,h} > C^S_{\alpha B} \), where \( C^S_{\alpha B} \) is the upper \( \alpha \)-percentile of the empirical distribution \( \{S^b_{d,h,j}\}_{j=1}^B \).

Although the procedure does not specify the number \( B \) of bootstrap replications to be carried out, in practice it is recommended to choose a number sufficiently large such that further increase does not substantially affect the critical values. Following Dwass (1957), MacKinnon (2007) pointed out that in addition, the number of bootstrap samples \( B \) must be such that the quantity \( \alpha(B + 1) \) is an integer, where \( \alpha \) is the level of the test. Moreover, as pointed out by Dufour & Khalaf (2001), the later requirement, together with the asymptotic pivotalness of the test statistics are necessary to get an exact bootstrap test.

The following theorem shows the uniform in bandwidth validity of the bootstrap method. This means that asymptotically, the probability law of the bootstrap test statistics \( HW^b_{d,h} \) and \( S^b_{d,h} \) given the data \( \{X_i, Y_i\}_{i=1}^n \) is respectively equivalent to the null asymptotic distribution of \( HW_{d,h} \) and \( S_{d,h} \) for almost all samples.

**Theorem 4.** Under the same conditions as in Lemma ??, then conditionally on the data and uniformly over \( h \in \mathcal{H}_n \),

\[
\sup_{h \in \mathcal{H}_n} \sup_{u \in \mathbb{R}} \left| \mathbb{P}(HW^b_{d,h} \leq u|\{X_i, Y_i\}_{i=1}^n) - \mathbb{P}(HW_{d,h} \leq u) \right| = o_p(1). 
\]

and

\[
\sup_{h \in \mathcal{H}_n} \sup_{u \in \mathbb{R}} \left| \mathbb{P}(S^b_{d,h} \leq u|\{X_i, Y_i\}_{i=1}^n) - \mathbb{P}(S_{d,h} \leq u) \right| = o_p(1). 
\]

(14)

(15)

In the next section we conduct Monte Carlo simulations to provide evidence on the behavior of our tests statistic in small and moderate samples, and compare our results with some existing tests.
5 Monte Carlo simulations

The setup is the one considered by Newey (1993), Tripathi & Kitamura (2003), Kitamura et al. (2004) and Otsu (2008):

\[ Y = \theta_1 + \theta_2 X + \nu, \text{ with } \theta_1 = \theta_2 = 1 \text{ and } \ln X \sim N(0,1). \]

For the error term \( \nu \), we consider two different situations:

- Heteroskedastic errors: \( \nu = \epsilon \sqrt{1 + .2X + .3X^2} \)
- Mixture errors:
  \[ \nu = \begin{cases} 
  \epsilon \sqrt{1 + .2X + .3X^2} & \text{with probability 0.9} \\
  \text{Cauchy}(0,1) & \text{with probability 0.1}
  \end{cases} \]

where \( \epsilon \sim N(0,1) \) and \( \epsilon \) is independent of \( X \). This setup is useful to compare our results with those of the above authors. We consider the SMD criterion with a gaussian kernel.

Our main focus in this setting is to examine the behavior of the specification test statistic under the null that the model is correctly specified, then observe its properties under a set of alternatives. Throughout this section, the null hypothesis is:

\[ H_0: \mathbb{E}[Y - \theta_1 - \theta_2 X|X] = 0 \text{ a.s. for some } (\theta_1, \theta_2) \]

The fixed bandwidth considered is \( d = 1 \), while the efficient bandwidth is taken as \( h_n = cn^{-1/5} \), where \( c = 0.5; 1; 1.5 \).

We examine the power performance of our tests when misspecification is present by evaluating their behavior under the following families of alternatives:

\[ H_1^A: Y = \theta_1 + \theta_2 X + sX^2 + \nu, \text{ with } s = 0.2, 0.3, 0.4 \]
\[ H_1^B: Y = \theta_1 + \theta_2 X + s\phi(X) + \nu, \text{ with } s = 3, 5, 7, \]

where \( \phi(\cdot) \) is the standard normal density function. The values of \( s \) are deviation from the null. The bigger the value of \( s \), the farther the alternative model is likely to be from the null. This is the same specification of alternatives used by Otsu (2008).

To investigate the small sample properties of our tests under the null and alternatives stated above, we compute 99 wild bootstrap statistics from 500 replications with sample sizes \( n = 50 \) and \( n = 100 \). At each replication, the critical values are estimated using the wild bootstrap procedure as described in the previous section. For the wild bootstrapping, the sample \( \{\omega_i, i = 1, \ldots, n\} \) is generated at each experiment via a two-point distribution defined by:
Note that this distribution has its first, second and third central moments all equal to one. As shown by Mammen 1992 for linear regression setups, this property is expected to provide better bootstrap approximations of the test statistic.

Our bootstrap simulation results are summarized in Table 1. The figures reported on the table are simulated rejection probabilities of the Hausman-type and the score-type bootstrap tests. The first panel of each model shows simulation results under the null. To analyze the sensitivity of our tests to the size of the bandwidth, three bandwidth coefficients \((c = 0.5, c = 1\) and \(c = 1.5\)) are considered here and provide different test patterns for each coefficient value. Although the power varies slightly with the bandwidth coefficient \(c\), the latter does not influence their overall behavior. It can be seen that in general the Score-type test have excellent empirical sizes at all sample and bandwidth sizes, with all rejection probabilities within the nominal size range of 5\%. As for the Hausman test, it slightly over rejects for relatively large bandwidths but have excellent sizes for all the other bandwidths. The power performance of both Hausman and score types bootstrap tests are also fairly good for the smaller sample \(n = 50\). For the sample size of \(n = 100\) both the Hausman-type and the score-type bootstrap tests have on average better size and power performance than the CEL, the SCEL and the ZHENG tests, for the family of alternatives \(H_{1}^{A}\) with heteroskedastic errors. To sum up, our tests statistics have reasonable sizes and power performance in our simulation experiments and are competitive with existing tests.

6 Conclusion

This paper has provided uniform in bandwidth specification tests for models defined by conditional moment restrictions. The Tests are built following either a Hausman (1978) approach or a score-type approach, and exploits the Lavergne & Patilea (2008) Smooth Minimum Distance criterion and estimators for CMR. Both test statistics are asymptotically equivalent, asymptotically chi-squared under the null hypothesis and are obtained uniformly within a wide range of bandwidths. We also analyze the asymptotic distribution of our tests under a fixed and a sequence of local alternatives and find that while having nontrivial power under local alternatives, they are powerful under fixed ones. Two bootstrap procedures are proposed to approximate the behavior of the tests statistics. We formally prove the validity of our bootstrap methods and use them to compute critical values of our tests in small samples. To our knowledge this is the first general bootstrap proposed to date for specification testing in nonlinear models defined by CMR. Both the test statistics and their bootstrap counterparts are simple to implement and a Monte Carlo simulation results shows that they perform well in small and moderate samples. Some directions to extend the proposed methods would be the
generalization of the testing procedure for the time series contexts, and the applications of our tests statistics to real economic data. We plan to explore these issues in further studies.

7 Appendix

7.1 Appendix A: Some useful formulas and notations

Let \( h \in \mathcal{H}_n \) and \( d \) be a fixed value of the bandwidth parameter. Define

\[
\triangle_{d,d} = \mathbb{E}[\nabla_\theta \tau(X_1, \theta_0) W^{-1/2}(X_1) W^{-1/2}(X_2) \text{Var}[g(Z_2, \theta_0)|X_2] W^{-1/2}(X_2) V_{d,d} \nabla'_\theta \tau(X_3, \theta_0) d^{-2d} K((X_1 - X_2)/d) K((X_2 - X_3)/d)],
\]

\[
V_d = \lim_{n \to \infty} V_{n,d} = \mathbb{E}[\nabla_\theta \tau(X_1, \theta_0) W^{-1/2}(X_1) W^{-1/2}(X_2) \nabla'_\theta \tau(X_2, \theta_0) h^{-q} d^{-2d} K((X_1 - X_2)/d)],
\]

and

\[
V_0 = \mathbb{E}[\nabla_\theta \mathbb{E}[g(Z_1, \theta_0)|X] \text{Var}^{-1}[g(Z_1, \theta_0)|X] \nabla'_\theta \mathbb{E}[g(Z_1, \theta_0)|X]];
\]

Let \( I = [I_p, -I_p] \) with \( I_p \) the \( p \times p \) identity matrix. Define \( \Omega = \begin{bmatrix} V_0^{-1} & 0 \\ 0 & V'_0 \end{bmatrix} \)

and \( \Sigma = \begin{bmatrix} \triangle_{d,d} & V_d \\ V_d & V_0 \end{bmatrix} \). Then, the asymptotic covariance matrix of \( \sqrt{n}(\hat{\theta}_{n,d} - \hat{\theta}_{n,h}) \) is given by \( Q_d = I \Omega \Sigma \Omega'^{-1} I' \).

A consistent estimator of \( Q_d \) is then defined by \( \hat{Q}_d = I \hat{\Omega} \hat{\Sigma} \hat{\Omega}' I' \). This expression is obtained by plugging-in consistent estimates of the matrices \( V_d, \triangle_{d,d} \) and \( V_0 \). Estimators for the later matrices can be respectively obtained using the following formulas:

\[
\frac{1}{2n(n-1)d^q} \sum_{1 \leq i \neq j \leq n} \nabla_\theta g(Z_i, \hat{\theta}_{n,d}) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) \nabla'_\theta g(Z_j, \hat{\theta}_{n,d}) K\left(\frac{X_i - X_j}{d}\right),
\]

(16)

\[
\frac{1}{n(n-1)(n-2)d^q} \sum_{1 \leq i \neq j \neq k \leq n} \nabla_\theta g(Z_i, \hat{\theta}_{n,d}) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) \text{Var}[g(Z_j, \hat{\theta}_{n,d}|X_j)]
\]

\[
\times W_n^{-1/2}(X_k) W_n^{-1/2}(X_j) \nabla'_\theta g(Z_j, \hat{\theta}_{n,d}) K\left(\frac{X_i - X_j}{d}\right) K\left(\frac{X_j - X_k}{d}\right),
\]

(17)

and

\[
\frac{1}{2n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \nabla_\theta g(Z_i, \hat{\theta}_{n,d}) \bar{W}_n^{-1/2}(X_i, \hat{\theta}_{n,h}) \bar{W}_n^{-1/2}(X_j, \hat{\theta}_{n,d}) \nabla'_\theta g(Z_j, \hat{\theta}_{n,d}) K\left(\frac{X_i - X_j}{h}\right)
\]

(18)
where \( \hat{\text{Var}}[g(Z_j, \theta|X_j)] \) and \( \hat{W}_n(X_j, \theta) \) are nonparametric consistent estimators of \( \text{Var}[g(Z_j, \theta|X_j)] \) and \( \text{Var}[g(Z_j, \theta|X_j)] f(X_j) \), respectively (see Equation (4) above).

We now introduce some empirical processes, functions and sets that are useful to derive our asymptotic results. Define: \( \mathcal{F}_n = \{ \phi_{n,h}(\cdot, \theta) : \theta \in \Theta, h \in [0, h_0] \} \) be the family of the measurable functions:

\[
\phi_{n,h}(z, \theta) = \mathbb{E}[\nabla_\theta \tau(X, \theta) W_n^{-1/2}(X) h^{-q} K((x - X)/h)] W_n^{-1/2}(x) g(z, \theta), \text{ for } h \in (0, h_0]
\]

Define: \( \mathbb{P}_n \phi_{n,h}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \phi_{n,h}(Z_i, \theta) \) and \( \mathbb{G}_n \phi_{n,h}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\phi_{n,h}(Z_i, \theta) - \mathbb{E}[\phi_{n,h}(Z_i, \theta)]] \)

\( \mathbb{P}_n \phi_{n,h}(\theta) \) and \( \mathbb{G}_n \phi_{n,h}(\theta) \) are sequences of empirical processes indexed by the families \( \mathcal{F}_n \).

It is useful to note that if Assumption 2(iii) is satisfied, then for any sequence of functions \( \psi_n \) that belong to the Sobolev space described in Assumption 6 the family of functions \( \{ x \mapsto \int \psi_n(x - uh) K(u) du : h \in [0, h_0] \} \) is uniformly Euclidean for an integrable envelope.

We also define the following non-random matrices:

\[
V_{n,h}(\theta) = H_{\theta,\theta} \mathbb{E} M_{n,h}(\theta, W_n)
\]

\[
= \mathbb{E}[\nabla_\theta \tau(X_1, \theta) W_n^{-1/2}(X_1) W_n^{-1/2}(X_2) \nabla_\theta \tau(X_2, \theta) h^{-q} K((X_1 - X_2)/h)]
\]

\[
+ \sum_{k=1}^{r} \mathbb{E}[H_{\theta,\theta} \tau_n^{(k)}(X_1, \theta) \tau_n^{(k)}(X_2, \theta) h^{-q} K((X_1 - X_2)/h)], \text{ for } h \in (0, h_0]
\]

\[
V_{n,0}(\theta) = \lim_{h \downarrow 0} V_{n,h}(\theta) = \mathbb{E}[\nabla_\theta \tau(X, \theta) W_n^{-1}(X) \nabla_\theta \tau(X, \theta) f(X)]
\]

\[
+ \sum_{k=1}^{r} \mathbb{E}[H_{\theta,\theta} \tau^{(k)}(X, \theta) W_n^{-1}(X) \tau^{(k)}(X, \theta) f(X)]
\]

The matrices \( V_{n,\theta}(\theta), h \in [0, h_0] \), are defined similarly with \( W(\cdot) \) in place of \( W_n(\cdot) \).

7.2 Appendix B: Preliminary Lemmas

In what follows, we establish some lemmas and we extend Theorem 2.3 and Theorem 3.2 of Lavergne & Patilea (2008). Those results are useful to prove our main results.
Lemma 7.1. Under Assumptions 1-4 and 7, and $H_0$, \( \sqrt{n}(\hat{\theta}_{n,h} - \theta_0) + V_{n,h}^{-1}(\theta_0)G_n\phi_{n,h}(\theta_0) = o_p(1) \), uniformly in $h \in \mathcal{H}_n$, and \( \{ G_n\phi_{n,h}(\theta_0) : h \in [0,h_0] \} \) weakly converges to a tight zero-mean Gaussian process.

Proof [ see Section 7.2 Lavergne & Patilea (2008)]

Lemma 7.2. Let Assumptions 1-7 and $H_0$ hold. Then, for all $\theta \in \Theta$, and uniformly over $h \in \mathcal{H}_n$,

\[
H_{\theta,\theta}M_{n,h}(\theta) - V_{n,h}(\theta) = o_p(1); \quad (19)
\]

\[
H_{\theta,\theta}M_{n,h}(\hat{\theta}_n) - V_{n,h}(\theta_0) = o_p(1); \quad (20)
\]

\[
\sqrt{n}\nabla_{\theta}M_{n,h}(\theta_0) - G\phi_{n,h}(\theta_0) = o_p(1); \quad (21)
\]

where $\hat{\theta}_n$ is any $\sqrt{n}$- consistent estimator of $\theta_0$.

Proof

Proof of Condition (19):

Denote:

\[
l(Z_1, Z_2; \theta, h) = \nabla_{\theta} g_n(Z_1, \theta) \nabla'_{\theta} g_n(Z_2, \theta) K((X_1 - X_2)/h) + \sum_{k=1}^r H_{\theta,\theta}g_{n,k}(Z_1, \theta) g_{n,k}(Z_2, \theta) K((X_1 - X_2)/h);
\]

We then have:

\[
H_{\theta,\theta}M_{n,h}(\theta, W_n) = U_n^2l(\cdot; \theta, h);
\]

By Assumption 6, the functions $\nabla_{\theta}g_{n,k}(\cdot, \theta)$ and $H_{\theta,\theta}g_{n,k}(\cdot, \theta)$ belong to the Sobolev space $H^s$ with $s > 3/2$. We can then use Assumption 4, Lemma 2.14 (ii) of Pakes and Pollard (1989) and Lemma 22 (ii) of Nolan and Pollard (1987) to conclude that the family of components of functions $l(Z_1, Z_2; \theta, h)$ and $E(l(Z_1, Z_2; \theta, h)$ indexed by $\theta$ and $h$ are uniformly euclidean for squared-integrable envelopes. By Corollary 7 of Sherman (1994a), we then have:

\[
\sup_{\theta \in \Theta, h > 0} ||H_{\theta,\theta}h^a(M_{n,h}(\theta, W_n) - EM_{n,h}(\theta, W_n))|| = sup_{\theta \in \Theta, h > 0} ||U_n^2[l(\cdot; \theta, h) - E(l(\cdot; \theta, h))|| = O_p(n^{-1/2})
\]

It follows that for all $\theta \in \Theta$,

\[
\sup_{h > 0} ||H_{\theta,\theta}h^a(M_{n,h}(\theta, W_n) - H_{\theta,\theta}EM_{n,h}(\theta, W_n))|| = O_p(n^{-1/2})
\]

Hence, \( \sup_{h \in \mathcal{H}_n} ||H_{\theta,\theta}h^a(M_{n,h}(\theta, W_n) - H_{\theta,\theta}EM_{n,h}(\theta, W_n))|| = O_P(n^{-1/2}). \)
Therefore:

$$||H_{\theta,\theta}M_{n,h}(\theta, W_n) - H_{\theta,\theta}EM_{n,h}(\theta, W_n)|| = O_p(\frac{1}{\sqrt{n}h^{2q}})$$

$$= o_p(1), \text{ uniformly over } h \in \mathcal{H}_n'.$$

That is:

$$||H_{\theta,\theta}M_{n,h}(\theta, W_n) - V_{n,h}(\theta)|| = o_p(1), \text{ uniformly over } h \in \mathcal{H}_n'.$$

**Proof of Condition (20):**

First, we have the following triangular inequality:

$$||H_{\theta,\theta}M_{n,h}(\hat{\theta}_n, W_n) - H_{\theta,\theta}EM_{n,h}(\hat{\theta}_n, W_n)|| = o_p(1), \text{ uniformly over } h \in \mathcal{H}_n'.$$

Next, we can replace $\theta$ by $\hat{\theta}_n$ in Condition (19) obtained above to get

$$||H_{\theta,\theta}M_{n,h}(\hat{\theta}_n) - V_{n,h}(\theta)|| = o_p(1), \text{ uniformly over } h \in \mathcal{H}_n'.$$

Thus, the first term of the right hand side of Inequality (22) converges in probability to zero.

As for the second term of Inequality (22), observe that under Assumptions 1-6, $\mathbb{E}H_{\theta,\theta}M_{n,h}(\theta)$ is a continuous function of $\theta$ and $h$. On the other hand, under $H_0$, the estimator $\hat{\theta}_n$ converges to $\theta_0$ in probability. Hence, by the Continuous Mapping Theorem, $\mathbb{E}H_{\theta,\theta}M_{n,h}(\hat{\theta}_n)$ converges to $\mathbb{E}H_{\theta,\theta}M_{n,h}(\theta_0)$. It then follows that: $||H_{\theta,\theta}EM_{n,h}(\hat{\theta}_n) - H_{\theta,\theta}EM_{n,h}(\theta_0)|| = o(1)$, under $H_0$

This achieves the proof of Condition (20).

**Proof of Condition (21):**

To see why Condition (21) holds, consider the family of the components of the functions
\[ q(Z_1, Z_2; \theta_0, h) = \nabla \theta g_n(Z_1, \theta_0) g_n(Z_2, \theta_0) K((X_1 - X_2)/h) - \mathbb{E}[\nabla \theta \tau_n(X, \theta_0) K((X - X_2)/h)] g_n(Z_1, \theta_0) \text{ indexed by } h > 0. \]

We then have: \( h^q \nabla \theta M_{n,h}(\theta_0, W_n) - h^q \mathbb{P} \phi_{n,h} = U_n^2 q(\cdot; \theta_0, h) \)

By an argument similar to the one used above, this family is uniformly Euclidean for a squared-integrable envelope. Moreover, for all \( h > 0 \)

\[
\mathbb{E}[q(Z_1, Z_2; \theta_0, h)|Z_2] = \mathbb{E}[q(Z_1, Z_2; \theta_0, h)|Z_1] \\
= \mathbb{E}[\nabla \theta \tau_n(X, \theta_0) K((X_1 - X)/h)] g_n(Z_1, \theta_0) - \mathbb{E}[\nabla \theta \tau_n(X, \theta_0) K((X_1 - X)/h)] g_n(Z_1, \theta_0) \\
= 0
\]

so that \( q(Z_1, Z_2; \theta_0, h) \) is degenerated. We can then apply Corollary 4 (ii) of Sherman (1994a) to get:

\[
\sup_{h>0} ||h^q \nabla \theta M_{n,h}(\theta_0, W_n) - h^q \mathbb{P} \phi_{n,h}(\theta_0)|| = O_P(1/n)
\]

Hence:

\[
\sup_{h>0} ||h^q \sqrt{n} \nabla \theta M_{n,h}(\theta_0, W_n) - h^q \mathbb{G} \phi_{n,h}(\theta_0)|| = O_P(n^{-1/2})
\]

It follows that:

\[
||\sqrt{n} \nabla \theta M_{n,h}(\theta_0, W_n) - \mathbb{G} \phi_{n,h}(\theta_0)|| = O_P(\inf_{h \in \mathcal{H}} \frac{1}{\sqrt{nh^2}}) \\
= o_P(1), \text{ uniformly over } h \in \mathcal{H}'. \]

**Lemma 7.3.** Under Assumptions 1-4 and 7, then conditionally on the sample and uniformly over \( h \in \mathcal{H}_n \),

(i) \( \sqrt{n}(\hat{\theta}_{n,h}^b - \hat{\theta}_{n,h}^a) \) and \( \sqrt{n}(\hat{\theta}_{n,d}^b - \hat{\theta}_{n,d}^a) \) have asymptotically the same distribution as \( \sqrt{n}(\hat{\theta}_{n,h} - \theta_0^a) \) and \( \sqrt{n}(\hat{\theta}_{n,d} - \theta_d^a) \), respectively. That is:

\[
\sup_{h \in \mathcal{H}_n} \sup_{u \in \mathbb{R}} |\mathbb{P}(\sqrt{n}(\hat{\theta}_{n,h}^b - \hat{\theta}_{n,h}) \leq u \{X_i, Y_i \}_{i=1}^n) - \mathbb{P}(\sqrt{n}(\hat{\theta}_{n,h} - \theta_0^a) \leq u)| = o_P(1),
\]

\[
\sup_{h \in \mathcal{H}_n} \sup_{u \in \mathbb{R}} |\mathbb{P}(\sqrt{n}(\hat{\theta}_{n,d}^b - \hat{\theta}_{n,d}) \leq u \{X_i, Y_i \}_{i=1}^n) - \mathbb{P}(\sqrt{n}(\hat{\theta}_{n,d} - \theta_d^a) \leq u)| = o_P(1).
\]

(ii) \( n(M_{n,h}^b(\hat{\theta}_{n,h}) - M_{n,h}^b(\hat{\theta}_{n,h})) \) and \( n(M_{n,d}^b(\hat{\theta}_{n,d}) - M_{n,d}^b(\hat{\theta}_{n,d})) \) have asymptotically the same distribution as \( n(M_{n,h}(\hat{\theta}_{n,h}) - M_{n,h}(\theta_0^a)) \) and \( n(M_{n,d}(\hat{\theta}_{n,d}) - M_{n,d}(\theta_d^a)) \), respectively.
\textbf{Proof}  [see section 7.2 Lavergne & Patilea 2008]

7.3 Appendix C: Main proofs

Proof of Lemma 1

Denote: \[ \Lambda_{d,h} = \sqrt{n}[H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d})]^{-1} \nabla_{\theta} M_{n,h}(\tilde{\theta}_{n,d}). \]
By first order expansion of \( \nabla_{\theta} M_{n,h}(\tilde{\theta}_{n,d}) \) around \( \theta_0 \) we have:

\[ \Lambda_{d,h} = [H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d})]^{-1} [\sqrt{n} \nabla_{\theta} M_{n,h}(\theta_0) + H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d}) \sqrt{n}(\tilde{\theta}_{n,d} - \theta_0)], \]

where \( \tilde{\theta}_{n,d} \) is between \( \tilde{\theta}_{n,d} \) and \( \theta_0 \).

Subtracting both sides of the above equation by \( \sqrt{n}(\hat{\theta}_{n,h} - \tilde{\theta}_{n,d}) \), we have:

\[ \Lambda_{d,h} - \sqrt{n}(\hat{\theta}_{n,h} - \tilde{\theta}_{n,d}) = \left[ [H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d})]^{-1} H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d}) - I \right] \sqrt{n}(\tilde{\theta}_{n,d} - \theta_0) + \left[ \sqrt{n}(\tilde{\theta}_{n,h} - \theta_0) + [H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d})]^{-1} \sqrt{n} \nabla_{\theta} M_{n,h}(\theta_0) \right] \]

(23)

By the \( \sqrt{n} \)-consistency of \( \tilde{\theta}_{n,d} \), \( \sqrt{n}(\tilde{\theta}_{n,d} - \theta_0) \) is bounded in probability. It follows by Lemma 7.2, Assumptions 3 and ?? and the continuity of matrix inversion that \( H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d}) = V_{nh}(\theta_0) + o_P(1) \) and \( [H_{\theta,\theta} M_{n,h}(\tilde{\theta}_{n,d})]^{-1} = V_{nh}^{-1}(\theta_0) + o_P(1) \) uniformly over \( h \in \mathcal{H}_n \), so that the first term of Equation (23) above converges to zero in probability, and Equation (23) reduces to:

\[ \Lambda_{d,h} - \sqrt{n}(\hat{\theta}_{n,h} - \tilde{\theta}_{n,d}) = - \left[ \sqrt{n}(\tilde{\theta}_{n,h} - \theta_0) + V_{nh}^{-1}(\theta_0) G_{\phi_{n,h}}(\theta_0) \right] + o_p(1) \]

(24)

Applying Lemma 7.1 to the right hand side of Equation (24), we have

\[ \Lambda_{d,h} = \sqrt{n}(\tilde{\theta}_{n,h} - \tilde{\theta}_{n,d}) + o_p(1). \]

(25)

We can now use Slutsky’s Lemma to conclude that

\[ \Lambda_{d,h}' Q_{d}^{-1} \Lambda_{d,h} = \sqrt{n}(\tilde{\theta}_{n,h} - \tilde{\theta}_{n,d})' Q_{d}^{-1} \sqrt{n}(\tilde{\theta}_{n,h} - \tilde{\theta}_{n,d}) + o_p(1). \]

(26)

That is, \( S_{d,h} = HW_{d,h} + o_p(1) \), uniformly over \( h \in \mathcal{H}_n \) for any fixed \( d \). \( \blacksquare \)
Proof of Theorem 1

Straightforward. We already know that \( \sqrt{n}(\hat{\theta}_{n,h} - \bar{\theta}_{n,d}) \stackrel{d}{\to} N(0, Q_d) \) uniformly over \( h \in \mathcal{H}_n \) for any fixed \( d \). If \( \hat{Q}_d \) is a consistent estimator of \( Q_d \), then \( HW_{d,h} = n(\hat{\theta}_{n,h} - \bar{\theta}_{n,d})' \hat{Q}_d^{-1}(\hat{\theta}_{n,h} - \bar{\theta}_{n,d}) \stackrel{d}{\to} \chi^2(p) \), where \( p = \text{rank}(Q_d) \). The asymptotic distribution of \( S_{d,h} \) then follows from that of \( HW_{d,h} \), given Lemma 1. ■

Proof of Theorem 2

Denote \( \delta = \theta^*_d - \theta^*_0 \). If \( \delta \neq 0 \), then \( \delta'Q_d^{-1}\delta \) is a strictly positive finite number.

Hence, uniformly over \( h \in \mathcal{H}_n \),

\[ n^{-1}HW_{d,h} = (\hat{\theta}_{n,h} - \bar{\theta}_{n,d})'(\hat{Q}_d^{-1}(\hat{\theta}_{n,h} - \bar{\theta}_{n,d}) \xrightarrow{P} \delta'Q_d^{-1}\delta > 0 \text{ as } n \to \infty; \]

which implies : \( HW_{d,h} \xrightarrow{P} +\infty \text{ as } n \to \infty \), and finally \( P_r[HW_{d,h} > C] \to 1 \) for any constant \( C \).

Likewise, by first order expansion, we have:

\[ \nabla_{\theta} M_{n,h}(\hat{\theta}_{n,h}) - \nabla_{\theta} M_{n,h}(\bar{\theta}_{n,h}) = H_{\theta,\theta} M_{n,h} (\bar{\theta}_{d,h}) (\bar{\theta}_{n,d} - \hat{\theta}_{n,h}) \]  

(27)

where \( \bar{\theta}_{d,h} \) is between \( \bar{\theta}_{n,d} \) and \( \hat{\theta}_{n,h} \).

Since \( \nabla_{\theta} M_{n,h}(\hat{\theta}_{n,h}) = 0 \), we then have

\[ \nabla_{\theta} M_{n,h}(\bar{\theta}_{n,h}) = H_{\theta,\theta} M_{n,h} (\bar{\theta}_{d,h}) (\bar{\theta}_{n,d} - \hat{\theta}_{n,h}) \]  

(28)

Denote:

\[ G(\theta) = \text{Plim} \nabla_{\theta} M_{n,h}(\theta) \quad \text{and} \quad \Omega(\theta) = \text{Plim} H_{\theta,\theta} M_{n,h}(\theta) \]  

(29)

Taking the probability limits uniformly over \( h \in \mathcal{H}_n \) both sides of the above Equation (28), and using Lemma 7.2, we get

\[ G(\theta^*_d) = \lim V_{n,h}(\bar{\theta}_{d,h}) \cdot (\theta^*_d - \theta^*_0) \]  

(30)

Since \( (\theta^*_d - \theta^*_0) \neq 0 \), and \( \lim V_{n,h}(\bar{\theta}_{d,h}) \) is nonsingular by Assumption??, we must have \( G(\theta^*_d) \neq 0 \). Also by Assumption??, this implies that \( [\Omega(\theta^*_d)]^{-1}G(\theta^*_d) \neq 0 \) so that uniformly over \( h \in \mathcal{H}_n \),

\[ n^{-1}S_{d,h} \xrightarrow{P} G(\theta^*_d)'[\Omega(\theta^*_d)Q_d\Omega(\theta^*_d)]^{-1}G(\theta^*_d) > 0 \text{ as } n \to \infty, \]  

(31)

Therefore, \( S_{d,h} \xrightarrow{P} \infty \), as \( n \to \infty \); and finally, \( P_r[S_{d,h} > C] \to 1 \) for any constant \( C \).■
Proof of Theorem 3

Under Assumptions 1-8, we have:

\[
\sqrt{n} (\tilde{\theta}_{n,d} - \theta_0) = -V_{n,d}^{-1}(\theta_0) [G_n \phi_{n,d}(\theta_0) - \eta_d] + o_p(1)
\]

\[
= -V_{n,d}^{-1}(\theta_0) G_n \phi_{n,d}(\theta_0) + V_{n,d}^{-1}(\theta_0) \eta_d + o_p(1)
\]

\[
\overset{d}{\rightarrow} N(V_{n,d}^{-1} \eta_d, V_{n,d}^{-1} \Delta_{d,d} V_{n,d}^{-1})
\]

We also have uniformly over \( h \in \mathcal{H}'_n \):

\[
\sqrt{n} (\hat{\theta}_{n,h} - \theta_0) = -V_{n,h}^{-1}(\theta_0) \left[ G_n \phi_{n,h}(\theta_0) - \eta_0 \right] + o_p(1)
\]

\[
= -V_{n,h}^{-1}(\theta_0) G_n \phi_{n,h}(\theta_0) + V_{n,h}^{-1}(\theta_0) \eta_0 + o_p(1)
\]

\[
\overset{d}{\rightarrow} N(V_0^{-1} \eta_0, V_0^{-1})
\]

Hence, \( \sqrt{n}(\tilde{\theta}_{n,d} - \hat{\theta}_{n,h}) \overset{d}{\rightarrow} N(\mu_d, Q_d) \). It follows that \( HW_{d,h} \overset{d}{\rightarrow} \chi^2(\mu'_d Q_d \mu_d, p) \) uniformly over \( h \in \mathcal{H}'_n \).

Likewise, we have

\[
\sqrt{n} \nabla_{\theta} M_{n,h}(\tilde{\theta}_{n,d}) = \sqrt{n} \nabla_{\theta} E_{M_{n,h}(\theta_0)} \left[ \nabla_{\theta} \tau(X, \theta_0) \text{Var}_0^{-1} [g(Z, \theta_0), X] \delta(X) \right] + \sqrt{n} \nabla_{\theta} (\tilde{\theta}_{n,d} - \theta_0) + o_p(1),
\]

\[
(32)
\]

Hence,

\[
\sqrt{n} [H_{\theta,0} M_{n,h}(\tilde{\theta}_{n,d})]^{-1} \nabla_{\theta} M_{n,h}(\tilde{\theta}_{n,d}) = V_{n,h}^{-1} G_n \phi_{n,h}(\theta_0) - V_{n,h}^{-1} E \left[ \nabla_{\theta} \tau(X, \theta_0) \text{Var}_0^{-1} [g(Z, \theta_0), X] \delta(X) \right] + \sqrt{n} (\tilde{\theta}_{n,d} - \theta_0) + o_p(1),
\]

\[
\overset{d}{\rightarrow} N(V_{d}^{-1} \eta_d - V_0^{-1} \eta_0, V_{d}^{-1} \Delta_{d,d} V_{d}^{-1} - V_0^{-1})
\]

\[
(33)
\]

It follows that \( S_{d,h} \overset{d}{\rightarrow} \chi^2(\mu'_d Q_d \mu_d, p) \), uniformly over \( h \in \mathcal{H}'_n \).

Proof of Theorem 4

Let \( \tilde{\theta}_{n,d}^b \) and \( \hat{\theta}_{n,h}^b \) denote the bootstrap versions of \( \tilde{\theta}_{n,d} \) and \( \hat{\theta}_{n,h} \). Denote by \( \theta_d^* \) and \( \theta_0^* \) the probability limits of the original corresponding estimators \( \tilde{\theta}_{n,d} \) and \( \hat{\theta}_{n,h} \). Note that
the values $\theta^*_d$ and $\theta^*_0$ are reduced to the true value $\theta_0$ if the model is correctly specify.

It is immediate from Lemma 7.3(i) that conditionally on the sample and uniformly over $h \in \mathcal{H}_n'$, $\sqrt{n}(\hat{\theta}_{n,d}^b - \theta_{n,d} + \hat{\theta}_{n,h} - \theta_{n,h})$ has asymptotically the same distribution as $\sqrt{n}(\hat{\theta}_{n,d} - \theta^*_d + \theta_0 - \hat{\theta}_{n,h})$.

Under $H_0$, we have $\theta^*_d = \theta_0$ and $\hat{Q}_d^b$ has asymptotically the same distribution as $\hat{Q}_d^b$ so that $HW_{d,h}^b$ is asymptotically $HW_{d,h}$ conditionally to the sample and therefore asymptotically $\chi^2(p)$.

Under $H_1$, $\sqrt{n}(\hat{\theta}_{n,d} - \theta^*_d + \theta_0 - \hat{\theta}_{n,h}) \xrightarrow{d} N(0, Q^*_d)$ and $\hat{Q}_d^b$ is asymptotically equivalent to $\hat{Q}_d^b$. Hence, conditionally on the sample and uniformly over $h \in \mathcal{H}_n'$, $HW_{d,h}^b$ is asymptotically equivalent to the statistic $\sqrt{n}(\hat{\theta}_{n,d} - \theta^*_d + \theta_0 - \hat{\theta}_{n,h})\hat{Q}_d^{-1}\sqrt{n}(\hat{\theta}_{n,d} - \theta^*_d + \theta_0 - \hat{\theta}_{n,h})$ which is asymptotically $\chi^2(p)$ as well.

Thus, conditionally on the sample and uniformly over $h \in \mathcal{H}_n'$, $HW_{d,h}^b$ has exactly the same asymptotic distribution as the null asymptotic distribution of $HW_{d,h}$.

Likewise, from Lemma 7.3(ii) and the continuity of the differentiation operator, it follows that conditionally on the sample and uniformly over $h \in \mathcal{H}_n'$,

$$\sqrt{n}(\nabla \theta E M_{n,h}(\hat{\theta}_{n,d}^b)) \xrightarrow{d} N(0, Q^*_d)$$

and $H_{\theta \theta} M_{n,h}(\hat{\theta}_{n,d}^b)$ is asymptotically equivalent to $H_{\theta \theta} M_{n,h}(\hat{\theta}_{n,d})$.

Under $H_0$, we have $\theta^*_d = \theta_0$. Hence, $\nabla \theta E M_{n,h}(\theta^*_0) = 0$ so that $S_{d,h}^b$ is asymptotically equivalent to $S_{d,h}$ conditionally to the sample, and therefore asymptotically $\chi^2(p)$.

Under $H_1$, $\sqrt{n}[H_{\theta \theta} M_{n,h}(\hat{\theta}_{n,d})]^{-1}\left(\nabla \theta E M_{n,h}(\hat{\theta}_{n,d}^b) - \nabla \theta E M_{n,h}(\theta^*_d)\right) \xrightarrow{d} N(0, Q^*_d)$ so that conditionally on the sample and uniformly over $h \in \mathcal{H}_n'$, $S_{d,h}^b$ is asymptotically equivalent to a quantity that is asymptotically $\chi^2(p)$ as well.

Thus, conditionally on the sample and uniformly over $h \in \mathcal{H}_n'$, $S_{d,h}^b$ has exactly the same asymptotic distribution as the null asymptotic distribution of $S_{d,h}$. ■
Bootstrap tests

Table 1: Percentage of rejection for the bootstrap tests (%)

<table>
<thead>
<tr>
<th>Models</th>
<th>$HW^b$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>$S^b$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=50</td>
<td>n=100</td>
<td>n=50</td>
<td>n=100</td>
<td>n=50</td>
<td>n=100</td>
<td>n=50</td>
<td>n=100</td>
<td>n=50</td>
</tr>
<tr>
<td>coeff c =</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Heterosk.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0$</td>
<td>5.3</td>
<td>4.4</td>
<td>6.3</td>
<td>4.8</td>
<td>5.1</td>
<td>6.7</td>
<td>4.67</td>
<td>5.3</td>
<td>4.5</td>
</tr>
<tr>
<td>$H_1^A$, $s = 0.2$</td>
<td>35.5</td>
<td>32.0</td>
<td>36.0</td>
<td>46.8</td>
<td>53.7</td>
<td>58.8</td>
<td>17.8</td>
<td>22.8</td>
<td>19.0</td>
</tr>
<tr>
<td>$H_1^A$, $s = 0.3$</td>
<td>53.3</td>
<td>65.5</td>
<td>72.0</td>
<td>76.0</td>
<td>80.0</td>
<td>82.4</td>
<td>26.3</td>
<td>34.0</td>
<td>28.2</td>
</tr>
<tr>
<td>$H_1^A$, $s = 0.4$</td>
<td>74.0</td>
<td>77.0</td>
<td>83.0</td>
<td>92.2</td>
<td>94.3</td>
<td>96.2</td>
<td>30.3</td>
<td>34.5</td>
<td>31.3</td>
</tr>
<tr>
<td>$H_1^B$, $s = 3$</td>
<td>8.0</td>
<td>11.0</td>
<td>10.4</td>
<td>10.0</td>
<td>9.8</td>
<td>16.6</td>
<td>7.0</td>
<td>9.0</td>
<td>10.0</td>
</tr>
<tr>
<td>$H_1^B$, $s = 5$</td>
<td>24.0</td>
<td>25.6</td>
<td>35.4</td>
<td>30.8</td>
<td>31.6</td>
<td>40.8</td>
<td>14.7</td>
<td>17.3</td>
<td>15.9</td>
</tr>
<tr>
<td>$H_1^B$, $s = 7$</td>
<td>38.0</td>
<td>40.0</td>
<td>52.6</td>
<td>53.4</td>
<td>58.4</td>
<td>67.4</td>
<td>23.7</td>
<td>28.0</td>
<td>26.5</td>
</tr>
<tr>
<td>Mixture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0$</td>
<td>4.6</td>
<td>5.6</td>
<td>6.0</td>
<td>4.0</td>
<td>5.0</td>
<td>8.0</td>
<td>5.2</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>$H_1^A$, $s = 0.2$</td>
<td>13.2</td>
<td>21.2</td>
<td>20.6</td>
<td>17.6</td>
<td>22.6</td>
<td>30.8</td>
<td>9.5</td>
<td>12.3</td>
<td>10.0</td>
</tr>
<tr>
<td>$H_1^A$, $s = 0.3$</td>
<td>25.4</td>
<td>36.2</td>
<td>37.2</td>
<td>34.2</td>
<td>42.8</td>
<td>44.8</td>
<td>15.5</td>
<td>16.8</td>
<td>13.0</td>
</tr>
<tr>
<td>$H_1^A$, $s = 0.4$</td>
<td>29.0</td>
<td>37.8</td>
<td>42.0</td>
<td>41.4</td>
<td>49.2</td>
<td>57.2</td>
<td>15.8</td>
<td>20.5</td>
<td>16.8</td>
</tr>
<tr>
<td>$H_1^B$, $s = 3$</td>
<td>8.6</td>
<td>9.0</td>
<td>12.0</td>
<td>10.8</td>
<td>9.8</td>
<td>16.8</td>
<td>9.3</td>
<td>10.0</td>
<td>11.2</td>
</tr>
<tr>
<td>$H_1^B$, $s = 5$</td>
<td>9.8</td>
<td>15.0</td>
<td>18.6</td>
<td>12.4</td>
<td>15.2</td>
<td>26.4</td>
<td>13.5</td>
<td>15.4</td>
<td>16.9</td>
</tr>
<tr>
<td>$H_1^B$, $s = 7$</td>
<td>11.0</td>
<td>18.2</td>
<td>24.4</td>
<td>21.8</td>
<td>26.2</td>
<td>34.2</td>
<td>20.7</td>
<td>23.0</td>
<td>24.5</td>
</tr>
</tbody>
</table>
References


