A Hausman Specification Test of Conditional Moment Restrictions*

Pascal Lavergne†　　Pierre E Nguimkeu‡

August 2011

Abstract

This paper addresses the issue of detecting misspecification of conditional moment restrictions (CMR). We propose a new Hausman-type test based on the comparison of an efficient estimator with an inefficient one, both derived by semiparametrically estimating the CMR using different bandwidths. The proposed test statistic is asymptotically chi-squared distributed under correct specification. We propose a general bootstrap procedure for computing critical values in small samples. The testing procedures are easy to implement and simulation results show that they perform well in small samples.

Keywords: Conditional Moment Restrictions, Hypothesis Testing, Smoothing Methods, Bootstrap.

JEL Classification: C52, C12, C14, C15.

*The second author gratefully acknowledges financial support from the SSHRC Canada. This paper was presented at the North American Summer Meeting of the Econometric Society 2011, the European Meeting of the Econometric society 2011, the Royal Economic Society Conference 2011, the Canadian Econometrics Study Group meeting 2010, and the African Econometric Society Conference 2010. We thank all conference participants, in particular, Russell Davidson, Arthur Lewbel, Victoria Zinde-Walsh, and Karim Abadir for useful comments.

†Toulouse School of Economics (GREMAQ); Manufacture des Tabacs, 21, Allée de Brienne, 31500 Toulouse France; Email: pascal.lavergne@univ-tlse1.fr

‡Corresponding author. Dept of Economics, Simon Fraser University; 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6; Email: npn1@sfu.ca, Webpage: www.sfu.ca/~npn1/

1 Introduction

This paper addresses the issue of detecting misspecification on models defined by conditional moment restrictions (CMR). Such models are pervasive in econometrics. The most popular example is the theory of dynamic optimizing agents with time separable utility where equilibrium conditions are typically stated and estimated in terms of martingale differences. 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 such models have been intensively investigated. 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 moments 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. 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). 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 bandwiths 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 estimation procedures rely on the crucial assumption that the Conditional Moment Restrictions under consideration are correctly specified. If the model is misspecified, the resulting estimators may have extremely different properties. A central issue for the practitioner is therefore to check the validity of these moment restrictions upon which estimation results crucially depend.

Some existing specification tests of CMR test 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). Dominguez and Lobato (2004), Delgado, Dominguez & Lavergne (2006) propose consistent specification tests based on a Cramer Von Mises criterion that are consistent against any alternative, but the asymptotic distribution of their tests statistic 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 smoothed empirical likelihood methods, that involve complex nonlinear optimization over many parameters, thus making the tests difficult to implement in practice.

This paper proposes a new practical procedure 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. The test is based on the distance between two SMD estimators: a consistent and asymptotically efficient one- indexed by a vanishing bandwidth - and a consistent but inefficient one - indexed by a fixed bandwidth. The test statistic is asymptotically chi-squared distributed under the null. We also propose bootstrap methods to approximate this test in small and moderate samples. The distributions and the validity of our bootstrap procedure are studied. Simulations show that the proposed specification test have good size and power performance in small and moderate samples.

The rest of the paper is organized as follows. In Section 2, we present the framework and the proposed test statistic. In Section 3, we discuss the asymptotic distribution and power properties. Bootstrap procedures to approximate the behavior of the test is proposed in Section 4. Section 5 reports Monte Carlo simulations results showing that our test 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

In this section, we describe our general framework for estimation and specification testing in CMR models, and explain the rationale for the proposed test. We use the following notations throughout. For a real valued function $l(\cdot)$, $\nabla_\theta l(\cdot)$ and $H_{\theta,\theta} l(\cdot)$ denote the p-column vector of first partial derivatives and the squared p matrix of second derivatives of $l(\cdot)$ with respect to the p-dimensional vector $\theta \in \mathbb{R}^p$. 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 θ.

Suppose we have a random sample of n independent observations $\{Z_i = (Y_i, X_i)\}_{i=1}^n$ on $Z = (Y', X')' \in \mathbb{R}^{s+q}$, $s \geq 1$, $q \geq 1$. X is distributed with Lebesgue density function $f(\cdot)$ while Y can 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 measurable functions of Z and of the p-dimensional parameter vector θ that belongs to a compact set $\Theta \subset \mathbb{R}^p$, $p \geq 1$. The conditional moment restrictions are defined by

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

Many econometric models are covered by this setup. In some contexts, the vector $g(Z, \theta)$ is the residual vector from some nonlinear multivariate regression. In others, $\mathbb{E}[g(Z, \theta_0)|X]$ is seen as the first order partial derivatives of some stochastic optimization problem.

Our test statistic use the Lavergne & Patilea (2008) smooth minimum distance (SMD) class of estimators for θ_0 characterized by (1). The typical SMD estimator obtains as the argument minimizing

$$M_{n,h}(\theta, W_n) = \frac{1}{2n(n-1)h^q} \sum_{1 \le i \ne j \le n} g'(Z_i, \theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij}^h \quad (2)$$

where $K_{ij}^h = K((X_i - X_j)/h)$, with $K(\cdot)$ a multivariate kernel, h a bandwidth parameter, and $W_n(\cdot)$ a sequence of $r \times r$ positive definite weighting matrices.

When the model is correctly specified, Lavergne and Patilea (2008) showed that a $\sqrt{n}$-consistent and asymptotically normal estimator can be obtained by minimizing (2) for $W_n(\cdot) = I_r$, the identity matrix, and a *fixed* bandwidth d, that is a bandwidth that does not depend on n. Moreover, a semiparametrically efficient SMD estimator $\widehat{\theta}_{n,h}$ follows from a two-step procedure where the second step uses a vanishing bandwidth h and a nonparametric estimator $\widehat{W}_n(\cdot)$ 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. For any preliminary consistent estimator $\check{\theta}$ of θ_0, we consider the estimator

$$\widehat{W}_n(x) = \frac{1}{n} \sum_{k=1}^{n} g(Z_k, \check{\theta}) g'(Z_k, \check{\theta}) b^{-q} K((x - X_k)/b), \quad (3)$$

where b is a bandwidth converging to zero. We note that a different kernel could also be used in the above estimator without affecting our results, as soon as this kernel satisfies the assumptions stated later.

However, a specification test is needed to check whether there exists a θ_0 such that the conditional moment restrictions (1) hold. Following an approach *à la* Hausman (1978), our proposed test is based on the distance between two SMD consistent estimators based on different bandwidths. More specifically, we focus in what follows on the comparison of an efficient estimator $\widehat{\theta}_{n,h}$ of θ_0, that uses a vanishing bandwidth h together with the estimated optimal weighting matrix (3), with a non efficient one $\widetilde{\theta}_{n,d}$, that uses a fixed bandwidth d but the same weighting matrix. Hence we define the test statictics $HW_{d,h}$ as

$$HW_{d,h} = n \left(\widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h} \right) \widehat{Q}_d^{-1} \left(\widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h} \right) \quad (4)$$

where $\widehat{Q}_d$ is a consistent estimator of Q_d, the asymptotic variance-covariance matrix of $\sqrt{n}(\widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h})$. When the model is correctly specified, both estimators are consistent for θ_0 so that their difference converge in probability to zero. The test statistic has then a simple chi-squared limiting distribution. In the presence of misspecification, the two estimators are expected to converge to different values in most cases, so that the distance between $\widehat{\theta}_{n,h}$ and $\widetilde{\theta}_{n,d}$ is nonzero even in large sample. Hence, significantly large values of $HW_{d,h}$ are regarded as evidence that the restrictions are not consistent with the data. Thus, the α-level test is $\mathbb{I}\left(HW_{d,h} > c_\alpha\right)$ where c_α is the $1-\alpha$ quantile of a χ^2_p distribution.

A practical drawback of our test is that in some instances the asymptotic variance of the estimator's differences could be singular, so that one should use a *modified* inverse, as proposed by Lutkepohl and Burda (1997), or a *regularized* inverse, as proposed by Dufour & Valery (2009). Our test statistic uses the optimal estimated weighting matrix for both estimators. Such a choice implies that $\widetilde{\theta}_{n,d}$ is computed in a supplementary step. Given that one already has at disposal a preliminary consistent estimator, this is easily done using a one quasi-Newton step.

3 Asymptotic Properties

We now provide regularity conditions under which the asymptotic properties of our specification test statistic is analyzed. In what follows, we denote by $\widehat{M}_{n,h}(\theta)$ the objective function that uses $\widehat{W}_n(\cdot)$ as defined by (3). Under correct specification, the objective function is then equivalent at first-order to the one using the true optimal weighting $\text{Var}[g(Z,\theta_0)|X=\cdot]f(\cdot)$, as shown by Lavergne and Patilea (2009).

Assumption 1. *(i) The parameter space* Θ *is compact.*
(ii) $\bar{\theta}_{n,h} = \arg\min_\Theta \mathbb{E}M_{n,h}(\theta)$ *is unique and belongs to* $\overset{\circ}{\Theta}$*, the interior of* Θ*.*

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

(iv) The density $f(\cdot)$ of X is bounded away from zero and infinity with bounded support D that can be written as finite unions and/or intersections of sets $\{x : p(x) \geq 0\}$, where $p(\cdot)$ is a polynomial function.

Let us define $\tau(x,\theta) = \mathbb{E}[g(Z,\theta)|X = x]$.

Assumption 3. *(i) 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 an envelope G with $\sup_{x\in\mathbb{R}^q} \mathbb{E}[G^8|X = x] < \infty$.
(iii) There exists $c > 0$ such that for all $\theta_1, \theta_2 \in \Theta$, $\mathbb{E}\|g(Z,\theta_1)-g(Z,\theta_2)\| \leq c\|\theta_1-\theta_2\|$
(iv) Let $\omega^2(\cdot,\theta) = \mathbb{E}[g(Z,\theta)g'(Z,\theta)|X = \cdot]$. Then, for all $\theta_1, \theta_2 \in \overset{\circ}{\Theta}$ and all $x \in \mathbb{R}^q$, $\|\omega^2(x,\theta_1) - \omega^2(x,\theta_2)\| \leq c\|\theta_1 - \theta_2\|^\nu$, for some $c > 0$ and $\nu > 2/3$.
(v) For any x, all second partial derivatives of $\tau(x,\cdot) = \mathbb{E}[g(Z,\cdot)|X = x]$ exist on $\overset{\circ}{\Theta}$. There exists a real valued function $H(\cdot)$ with $\mathbb{E}H^4 < \infty$ and some constant $a \in (0,1]$ such that:
$\|H_{\theta,\theta}\tau^{(k)}(X,\theta_1) - H_{\theta,\theta}\tau^{(k)}(X,\theta_2)\| \leq H(Z)\|\theta_1 - \theta_2\|^a$, $\forall\ \theta_1, \theta_2 \in \overset{\circ}{\Theta}$, $k = 1,\ldots,r$.
(vi) The components of $\nabla_\theta\tau(\cdot,\theta_1)f(\cdot)$ and of $\mathbb{E}\left[g(Z,\theta_1)g'(Z,\theta_2)|X = \cdot\right] f(\cdot)$, $\theta_1, \theta_2 \in \overset{\circ}{\Theta}$, are uniformly bounded in $L^1 \cap L^2$ and are continuous in $\theta_1, \theta_2 \in \overset{\circ}{\Theta}$.

Assumption 4. *When(1) holds, (i) $\mathbb{E}\left[\nabla_\theta\tau(X,\theta_0)\nabla'_\theta\tau(X,\theta_0)\right]$ is non singular. (ii) Each of the entries of $\nabla_\theta\tau(\cdot,\theta_0)f(\cdot)$, $\mathrm{H}_{\theta,\theta}\tau^{(k)}(\cdot,\theta_0)f(\cdot)$, $1 \leq k \leq r$ and $H(\cdot)f(\cdot)$ is Hölder continuous on D, with possibly different exponents.*

Under correct specification, that is if the conditional moment restrictions (1) hold for a unique θ_0, then $\bar{\theta}_{n,h} = \theta_0$ in Assumption 1. 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 $K(\cdot)$ is useful to establish consistency of SMD estimators (see Lavergne & Patilea 2008). Assumption 2 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 garantees in particular that $\mathbb{E}M_{n,h}(\theta)$ is a continuous function with respect to both θ and h, and that under H_0 the second step estimator $\widehat{\theta}_{n,h}$ is asymptotically efficient. Twice differentiability of $g(z,\cdot)$ is not strictly needed for the construction of our Hausman test statistic. This allows the specification test to apply to a wider variety of models including conditional quantile restrictions. Assumption 4 is needed only when studying the tests' behavior under correct specification of (1). Part (i) is a standard local identification condition.

We first sum up the main properties of the SMD estimators that follow from results by Lavergne and Patilea (2008). Let $\mathcal{H}_n = \{1/\ln(n+1) \geq h > 0 : nh^{4q/\alpha} \geq C\}$ where $C > 0$ and $\alpha \in (0,1)$ are arbitrary constants. Under (1), and for any fixed d, $\sqrt{n}\left(\widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h}\right)$, considered as a process indexed by $h \in \mathcal{H}_n$, converges in distribution to a tight process whose marginals are zero-mean normal with covariance function given by Q_d. The definition of Q_d, as well as its estimator $\widehat{Q}_d$, is given in Section 7.

Hence, when the model is correctly specified, the test statistic has the asymptotic behavior stated below.

Theorem 1. *Let Assumptions 1-4 hold. Then under (1) and for any fixed d,* $HW_{d,h}$ *converge in distribution to a* χ^2_p *uniformly over* $h \in \mathcal{H}_n$.

In the presence of misspecification, the population conditional moment $\mathbb{E}[g(Z,\theta)|X]$ is different from zero for any value of the parameter θ. In this case, SMD estimators $\widetilde{\theta}_{n,d}$ and $\widehat{\theta}_{n,h}$ would typically converge to two different limits. Denote $\bar{\theta}_d$ and $\bar{\theta}_0$ the probability limits of $\widetilde{\theta}_{n,d}$ and $\widehat{\theta}_{n,h}$ when the model is misspecified, and by $\overline{Q}_d$ the probability limit of $\widehat{Q}_d$.

Theorem 2. *Let Assumptions 1-3 hold. Then uniformly over* $h \in \mathcal{H}_n$ $HW_{d,h} \xrightarrow{p} +\infty$ *provided* $\overline{Q}_d^{-1}\left(\bar{\theta}_d - \bar{\theta}_0\right) \neq 0$.

The above result makes clear that the test might not be consistent in some circumstances, and in particular, if the two estimators have the same probability limit.

4 Bootstrap Tests

Bootstrapping is popular to approximate the distribution of statistics when asymptotics may not reflect accurately their behavior in small or moderate samples. For testing specification (1), application of bootstrap would require to generate resamples with the same values of X, but new observations for Y that fulfill the moment restrictions. This can be done easily in simple cases, e.g. wild bootstrap in regression models, and has been shown to give reliable approximations in many situations. In general however, generating bootstrap samples may be difficult or even infeasible: in simultaneous equations systems that are nonlinear in the variables Y, a reduced form may not be available or unique. We here propose a simple method that allows to circumvent these difficulties if they appear, that applies generally and is easy to implement. This method has been proposed by Jin, Ying and Wei (2001) and Bose and Chatterjee (2003), see also Chatterjee and Bose (2005) for a similar method applied to Z-estimators and Chen and Pouzo (2009) for sieve minimum distance estimators. However, they impose conditions that do not hold in our context. More crucially, they do not investigate the use of this method for specification testing.

Instead of resampling observations, we perturb the objective function and recompute our test statistic using this perturbed objective function. Consider n independent identical copies $w_i, i = 1, \ldots n$, of a known positive random variable w with $\mathbb{E}(w) = \text{Var}(w) = 1$ and $\mathbb{E}w^4 < \infty$. Define the new perturbed criterion as

$$M^*_{n,h}(\theta) = \frac{1}{2n(n-1)} \sum_{1 \le i \ne j \le n} w_i w_j g'(Z_i, \theta) \widehat{W}_n^{-1/2}(X_i) \widehat{W}_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij} \,.$$

We can then compute new SMD estimators based on the perturbed objective function. Since the w_i, $i = 1, \ldots n$, are independent of the original sample, it is easy to see that under the above conditions $\mathbb{E}[wg(Z,\theta)|X] = \mathbb{E}[g(Z,\theta)|X]$ so that the perturbed function still fulfills the moment restrictions whenever the original function does. With the new criterion, we repeat the optimization process by estimating $\widetilde{\theta}^*_{n,d}$, the bootstrap SMD estimator with fixed bandwidth d and $\widehat{\theta}^*_{n,h}$, the efficient one with vanishing bandwidth h. In practice, one could simply use a Newton-Raphson step from the original estimators to update to the new estimators. We can then compute the bootstrap version of our test statistic by

$$HW^*_{d,h} = n\left((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h})\right)' \widehat{Q}^{*,-1}_d \left((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h})\right) ,$$

where Q^*_d is the bootstrap counterpart of Q_d and $\widetilde{\theta}_{n,d}$ and $\widetilde{\theta}_{n,h}$ are the original non-bootstrap SMD estimators. The process is repeated a large number of times, say B, to obtain an empirical distribution of the B bootstrap test statistics $\{HW^*_{d,h,j}\}_{j=1}^B$. This bootstrap empirical distribution is then used to approximate the distribution of the test statistic $HW_{d,h}$ under correct specification, allowing to calculate the critical values empirically. Typically, one rejects H_0 at α level if $HW_{d,h} > c^H_{\alpha,B}$, where $c^H_{\alpha,B}$ is the upper α-percentile of the empirical distribution $\{HW^*_{d,h,j}\}_{j=1}^B$.

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 α 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.

Theorem 3. *Under the assumption of Theorem 2, then conditionally on the sample*
(i) Under H_0, $\sup_{h\in\mathcal{H}_n} \sup_{u\in\mathbb{R}} \left|\mathbb{P}(HW^*_{d,h} \leq u|\{Z_i\}_{i=1}^n) - \mathbb{P}(HW_{d,h} \leq u)\right| = o_p(1)$,
(ii) When H_0 *does not hold,* $HW^*_{d,h} = o_p(n)$ *uniformly over* $h \in \mathcal{H}_n$

Since $HW_{d,h}$ diverges at rate n under the alternative as given in Theorem 2, the second part of the theorem implies that $\mathbb{P}[HW_{d,h} > HW^*_{d,h}] \xrightarrow{p} 1$ when $n \to \infty$, which suffices to obtain a consistent test.

5 Monte Carlo simulations

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

$Y = \theta_1 + \theta_2 X + \nu$, with $\theta_1 = \theta_2 = 1$ and $\ln X \sim N(0, 1)$.

For the error term ν, we consider three different situations:

- Homoskedastic errors : $\nu = \epsilon$.

- 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 ϵ 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: $\mathbb{E}[Y - \theta_1 - \theta_2 X|X] = 0$ a.s. for some (θ_1, θ_2)

The fixed bandwidth considered is $d = 1$, while the efficient bandwidth is taken as $h_n = n^{-1/5}$

We examine the power performance of our tests when misspecification is present by evaluating their behavior under the following data generating processes (DGP):

$H_1^A : Y = 1 + X + sX^2 + \nu$, with $s = 0, 0.2, 0.3, 0.4$

$H_1^B : Y = 1 + X + s\phi(X) + \nu$, with $s = 0, 3, 5, 7$,

where $\phi(\cdot)$ is the standard normal density function. The values of s are deviation from the null. For $s = 0$, the model being tested is correctly specified. The bigger the value of s, the farther the data generating process is likely to be from the null. This is the same specification of DGPs used by Otsu (2008).

Table 1 summarizes both our general and bootstrap tests statistics results and also reports results for other specification tests. All the existing tests compared here with our general test, denoted HW, are obtain from 1000 replications at a nominal significance level of 5% with a sample size of $n = 100$. The figures reported on the table are simulated rejection probabilities. The first row of each model reports simulation results under the null, thus showing the size of each test. While our general test slightly over-rejects the null, it has a very good power under the alternatives, especially for model H_1^A. Our results are compared to four other methods featured in the Otsu (2008) simulation study: the method of conditional empirical likelihood (CEL) proposed by Kitamura et al. (2004), the method of smoothed empirical likelihood (SCEL) by Otsu (2008), the Zheng's (1998) test, and the Ramsey Regression Equation Specification Error Test (RESET). Our general test has on average better power performance than the CEL, the SCEL, the ZHENG and the RESET tests, for the family of alternatives H_1^A. For the family of alternatives H_1^B, the power performance of our general test is also good and compete with others.

For our bootstrap test, denoted HW^b, we computed 199 bootstrap statistics from 1000 replications with sample size $n = 100$. At each replication, critical values at 5% significance are estimated by taking the 95^th upper percentile of the distribution of bootstrap values as explained in the bootstrap procedure presented in section 4. For the wild bootstrapping, the sample $\{\omega_i, i = 1, \ldots, n\}$ is generated at each experiment via a two-point distribution defined by:

$$\mathbb{P}\left[\omega_i = \frac{3-\sqrt{5}}{2}\right] = 1 - \mathbb{P}\left[\omega_i = \frac{3+\sqrt{5}}{2}\right] = \frac{5+\sqrt{5}}{10}$$

Note that this distribution has its first, second and third central moment 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. As reported in the first column of the table our bootstrap test has a very good empirical size since all rejection probabilities are within the nominal range of 5%. Moreover, the empirical size performance of our bootstrap test is superior to all the other tests. The power performance of the bootstrap test is also fairly good, though less good than our asymptotic test and other tests. This feature is however expected since a gain in size is often traded with a relative loss in power in the bootstrap test due to its conservative nature. To sum up, our general test statistic has very good power performance in our simulation experiments and are competitive with existing tests. Moreover, the superiority of the empirical size performance of our bootstrap shows that our bootstrap test can properly handle small sample size models.

6 Conclusion

This paper has provided a new specification test for models defined by conditional moment restrictions. The Test is built following a Hausman (1978) approach and exploits the Lavergne & Patilea (2008) Smooth Minimum Distance estimators for CMR. The test statistic is asymptotically chi-squared under the null hypothesis and diverges under the alternative, uniformly within a wide range of bandwidths. A bootstrap procedure is proposed to approximate the behavior of the test statistic in small samples. We formally prove the validity of our bootstrap method and use them to compute critical values of our tests. Both the test statistic and its bootstrap counterpart is simple to implement and a Monte Carlo simulation study shows that they perform well in small and moderate samples and are competitive with existing tests. Moreover, the test require weaker mathematical assumptions about the estimating function so that it readily applies to a wider variety of models. Some directions to extend the proposed methods would be the generalization of the testing procedure for the time series context. We plan to explore these issues in further studies.

7 Technical material

In what follows, we denote $\check{\theta}$ any preliminary estimator of θ_0 and $\bar{\theta}$ the probability limit of $\check{\theta}$, which coincides with θ_0 when the model is correctly specified. Let $W_n(x,\check{\theta}) = \mathbb{E}[\widehat{W}_n(x,\check{\theta})]$, where $\widehat{W}_n(x,\check{\theta})$ $\big($also denoted $\widehat{W}(x)$, for simplicity$\big)$ is the estimator of the optimal weighting matrix given by (3) and denote $W_n(x) = W_n(x,\bar{\theta})$.

The sequence $W_n(x)$ is a non-random process indexed by the bandwith $b \in \mathcal{H}_n$ and its pointwise limit is denoted $W(x) = \lim W_n(x)$. Unless otherwise specified, we denote $\widehat{M}_{n,h}(\theta)$ $\big($respectively, $M_{n,h}(\theta)\big)$ the objective function given in (2) with the weighting matrix $\widehat{W}_n(x)$ (respectively, $W_n(x)$). Note that $\widehat{M}_{n,h}(\theta)$ and $M_{n,h}(\theta)$ are processes indexed by both the bandwidths h and b.

7.1 SMD estimation

Let $\phi_d(z,\theta) = \mathbb{E}[\nabla_\theta \tau(X,\theta) W^{-1/2}(X) d^{-q} K((x-X)/d)] W^{-1/2}(x) g(z,\theta)$, $\phi_0(z,\theta) = \nabla_\theta \tau(x,\theta) W^{-1}(x) f(x) g(z,\theta)$, and $\mathbb{G}_n \phi(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\phi(Z_i,\theta) - \mathbb{E}\phi(Z_i,\theta)]$. Define

$$
\begin{aligned}
V_d &= \mathbb{E}[\nabla_\theta \tau(X_1,\theta_0) W^{-1/2}(X_1) W^{-1/2}(X_2) \nabla'_\theta \tau(X_2,\theta_0) d^{-q} K((X_1 - X_2)/d)] \\
V_0 &= \mathbb{E}\big[\nabla_\theta \tau(X,\theta_0) W^{-1}(X) [\nabla'_\theta \tau(X,\theta_0) f(X)\big] .
\end{aligned}
$$

Lemma 7.1. *Under Assumptions 1-4 and (1), then (i)* $\sqrt{n}\big(\widehat{\theta}_{n,h} - \theta_0\big) + V_0^{-1} \mathbb{G}_n \phi_0(\theta_0) = o_p(1)$*, uniformly in* $h, b \in \mathcal{H}_n$*, where* $\mathbb{G}_n \phi_{n,h}(\theta_0)$ *weakly converges to a* $N(0, V_0)$*.*

(ii) $\sqrt{n}\big(\widetilde{\theta}_{n,d} - \theta_0\big) + V_d^{-1} \mathbb{G}_n \phi_d(\theta_0) = o_p(1)$ *uniformly in* $b \in \mathcal{H}_n$ *for any fixed* d*, where* $\mathbb{G}_n \phi_d(\theta_0)$ *weakly converges to a* $N(0, \Delta_d)$*, with*

$$
\begin{aligned}
\Delta_{d,d} = {} & \mathbb{E}[\nabla_\theta \tau(X_1,\theta_0) W^{-1/2}(X_1) W^{-1/2}(X_3) \nabla'_\theta \tau(X_3,\theta_0) f^{-1}(X_2) \\
& d^{-2q} K((X_1 - X_2)/d) K((X_2 - X_3)/d)] .
\end{aligned}
$$

(iii) $\sqrt{n}\big(\tilde{\theta}_{n,d} - \widehat{\theta}_{n,h}\big)$ *weakly converges to a* $N(0, Q_d)$ *for any fixed* d *and uniformly in* $h, b \in \mathcal{H}_n$*, where* $Q_d = V_d^{-1} \Delta_d V_d^{-1} - V_0^{-1}$*.*

Proof. Part (i) follows directly from Section 5.2 of Lavergne & Patilea (2008). Part (ii) follows similarly by noticing that their condition (2.7) also holds for $\widehat{M}_{n,d}(\theta)$, where d is a fixed bandwidth. Part (iii) follows from (i) and (ii). □

An estimator of Q_d is given by $\widehat{Q}_d = \widehat{V}_d^{-1} \widehat{\Delta}_d \widehat{V}_d^{-1} - \widehat{V}_0^{-1}$ where the respective estimators of V_d, V_0, and $\Delta_{d,d}$ are

$$
\frac{1}{n(n-1)} \sum_{i \neq j} \nabla_\theta g(Z_i, \tilde{\theta}_{n,d}) \widehat{W}_n^{-1/2}(X_i) \widehat{W}_n^{-1/2}(X_j) \nabla'_\theta g(Z_j, \tilde{\theta}_{n,d}) d^{-q} K\big(\frac{X_i - X_j}{d}\big) ,
$$

$$\frac{1}{n}\sum_i \nabla_\theta g(Z_i,\widehat{\theta}_{n,h})\widehat{W}_n^{-1}(X_i)f_n(X_i)\nabla'_\theta g(Z_i,\widehat{\theta}_{n,h}) \text{ and}$$

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq k, j\neq k}\nabla_\theta g(Z_i,\tilde{\theta}_{n,d})\widehat{W}_n^{-1/2}(X_i)\widehat{W}_n^{-1/2}(X_k)\nabla'_\theta g(Z_k,\tilde{\theta}_{n,d})f_n^{-1}(X_j)$$
$$d^{-2q}K\big(\tfrac{X_i-X_j}{d}\big)K\big(\tfrac{X_j-X_k}{d}\big),$$

where $f_n(X_i)=\frac{1}{n-1}\sum_{j\neq i}h^{-q}K((X_i-X_j)/h))$ is the leave-one-out kernel estimator of $f(X_i)$

Lemma 7.2. *Let $A,B\in\mathbb{R}^{n\times p}$ be random matrices such that $\mathbb{E}\|A\|<\infty$, $\mathbb{E}\|B\|<\infty$. Suppose $\mathbb{E}(A'B)$, $\mathbb{E}(B'A)$, and $\mathbb{E}(B'B)$ are non-singular matrices.*
Then $\mathbb{E}^{-1}(B'A)\mathbb{E}(A'A)\mathbb{E}^{-1}(A'B)-\mathbb{E}^{-1}(B'B)$ is positive semidefinite, with equality iff $B=A\mathbb{E}^{-1}(B'A)\mathbb{E}(B'B)$.

Proof. Consider $C=A\mathbb{E}^{-1}(B'A)-B\mathbb{E}^{-1}(B'B)\in\mathbb{R}^{n\times p}$. Then

$$\mathbb{E}[C'C]=\mathbb{E}^{-1}(B'A)\mathbb{E}(A'A)\mathbb{E}^{-1}(A'B)-\mathbb{E}^{-1}(B'B)$$

is positive semidefinite by definition, as the expectation of a matrix product of the form $C'C$, and is zero if and only if $C=0$. Conclude by noticing that $C=0$ is equivalent to $B=A\mathbb{E}^{-1}(B'A)\mathbb{E}(B'B)$ □

Lemma 7.3. *Let Assumptions 1-4 and (1) hold. Then, uniformly in $h,b\in\mathcal{H}_n$ and for any fixed d,*
(i) $\widehat{Q}_d=Q_d+o_p(1)$
(ii) Q_d is positive semidefinite.

Proof. For part (i), we only need to prove that the matrices $\widehat{V}_d$, $\widehat{\Delta}_d$ and $\widehat{V}_0$ converge in probability to V_d, Δ_d and V_0 respectively, and use the continuous mapping theorem to conclude. The convergence results for those matrices can be found in Section 5.2 of Lavergne & Patilea (2008).
For part (ii), apply Lemma 7.2 with $A=\mathbb{E}\big[W^{-1/2}(X_2)\nabla'_\theta\tau(X_2,\theta_0)d^{-q}K((X-X_2)/d)\big]\,f^{-1/2}(X)$ and $B=W^{-1/2}(X)\nabla'_\theta\tau(X,\theta_0)f^{1/2}(X)$. The desired conclusion then follows. □

7.2 Asymptotic behavior of the tests

Proof of Theorem 1

The result follows from Lemmas 7.1 and 7.3.□

Proof of Theorem 2

Denote $\delta = \bar{\theta}_d - \bar{\theta}_0$. If $\overline{Q}_d^{-1}\delta \neq 0$, then $\delta'\overline{Q}_d^{-1}\delta$ is a strictly positive finite number. Hence, uniformly over $h, b \in \mathcal{H}_n$, $n^{-1}HW_{d,h} = (\widehat{\theta}_{n,h} - \widetilde{\theta}_{n,d})'\widehat{Q}_d^{-1}(\widehat{\theta}_{n,h} - \widetilde{\theta}_{n,d}) \xrightarrow{p} \delta'\overline{Q}_d^{-1}\delta > 0$ as $n \to \infty$. It follows that $HW_{d,h} \xrightarrow{P} +\infty$ as $n \to \infty$. □

Lemma 7.4. *Let Assumptions 1-3 hold. Then uniformly over $h, b \in \mathcal{H}_n$,*

$$\sup_{\theta\in\Theta}\left|\widehat{M}_{n,h}(\theta) - M_{n,h}(\theta)\right| = o_p(1), \tag{5}$$

Proof.

The proof proceeds in two steps.

Step 1 is to show that for any $\bar{\theta} \in \Theta$, $\sup_{x\in\mathbb{R}^q}\left\|\widehat{W}_n(x,\theta) - W_n(x,\bar{\theta})\right\| = o_p(1)$ uniformly over $b \in \mathcal{H}_n$ and θ in an $o(1)$ neighborhood of $\bar{\theta}$. For this purpose, we apply a useful result given by Theorem 2 of Einmahl & Mason (2005) that establishes that $\sup_{x\in\mathbb{R}^q}\left\|\widehat{W}_n(x,\theta) - W_n(x,\theta)\right\| = o_p(1)$ uniformly in $\theta \in \Theta$ and over $b \in \mathcal{H}_n$. This result is true in this framework provided their condition (1.7) on the continuity of the density $f(\cdot)$ is replaced by the condition of a bounded density as given by our Assumption 2(iv). On the other hand, by our Assumption 3(iv) we have

$$\sup_{x\in\mathbb{R}^q}\left\|W_n(x,\theta) - W_n(x,\bar{\theta})]\right\| \leq c\|\theta - \bar{\theta}\|^{\nu}\|\mathbb{E}[b^{-q}K((X-x)/b)]\| \leq C\|\theta - \bar{\theta}\|^{\nu},$$

for some constant $C > 0$. It then follows that for any $\bar{\theta}$,

$$\begin{aligned}\sup_{x\in\mathbb{R}^q}\left\|\widehat{W}_n(x,\theta) - W_n(x,\bar{\theta})\right\| &\leq \sup_{x\in\mathbb{R}^q}\left\|\widehat{W}_n(x,\theta) - W_n(x,\theta)\right\| + \sup_{x\in\mathbb{R}^q}\left\|W_n(x,\theta) - W_n(x,\bar{\theta})\right\| \\ &\leq o_p(1) + C\|\theta - \bar{\theta}\|^{\nu}\end{aligned}$$

Hence, $\sup_{x\in\mathbb{R}^q}\left\|\widehat{W}_n(x,\theta) - W_n(x,\bar{\theta})\right\| = o_p(1)$ uniformly over θ in an $o(1)$ neighborhood of $\bar{\theta}$. I then follows that for any preliminary estimator $\check{\theta}$, of θ_0, $\sup_{x\in\mathbb{R}^q}\left\|\widehat{W}_n(x,\check{\theta}) - W_n(x)\right\| =$

$o_p(1)$.

Step 2 uses the result of *Step 1* to show Condition (5). For this purpose, we can write $\widehat{M}_{n,h}(\theta) - M_{n,h}(\theta) = M_{1n} + M_{2n}$, where $M_{1n} = M_{1n}(\theta, h, b)$ and $M_{2n} = M_{2n}(\theta, h, b)$ are given by

$$M_{1n} = \frac{h^{-q}}{2n(n-1)} \sum_{i \neq j} g'(Z_i, \theta) \widehat{W}_n^{-1/2}(X_i, \check{\theta})[\widehat{W}_n^{-1/2}(X_j, \check{\theta}) - W_n^{-1/2}(X_j)] g(Z_j, \theta) K_{ij}$$

$$M_{2n} = \frac{h^{-q}}{2n(n-1)} \sum_{i \neq j} g'(Z_i, \theta)[\widehat{W}_n^{-1/2}(X_i, \check{\theta}) - W_n^{-1/2}(X_i, \check{\theta})] W_n^{-1/2}(X_j)] g(Z_j, \theta) K_{ij}$$

Let A and B be any two positive definite matrices. Since the euclidean matrix norm $\| \cdot \|$ is unitarily invariant, then by Theorem 6.2 of Higham (2008) we have $\|A^{1/2} - B^{1/2}\| \leq \dfrac{1}{\lambda_{\min}(A)^{1/2} + \lambda_{\min}(B)^{1/2}} \|A - B\|$. If we write $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, it then follows that

$$\|A^{-1/2} - B^{-1/2}\| \leq \frac{1}{\lambda_{\min}(A)^{-1/2} + \lambda_{\min}(B)^{-1/2}} \|A^{-1}\| \|B^{-1}\| \|A - B\|$$

Our Assumption 3(iii) together with Assumption 1(i) and *step 1* guarantee that both $\widehat{W}_n^{-s}(x, \cdot)$ and $W_n^{-s}(x, \cdot)$, $s = 1, \frac{1}{2}$, and their eigenvalues are uniformly bounded. Hence, by the above inequality, there exists some constant $C_1 > 0$ such that

$$\sup_{x \in \mathbb{R}^q} \left\| \widehat{W}_n^{-1/2}(x, \check{\theta})[\widehat{W}_n^{-1/2}(x, \check{\theta}) - W_n^{-1/2}(x)] \right\| \leq C_1 \sup_{x \in \mathbb{R}^q} \left\| \widehat{W}_n(x, \check{\theta}) - W_n(x) \right\|,$$

Thus, uniformly over $h, b \in \mathcal{H}_n$,

$$\begin{aligned} \|M_{1n}\| &\leq \frac{C_1}{2n(n-1)h^q} \sum_{i \neq j} \|g(Z_i, \theta)\| \|g(Z_j, \theta)\| K_{ij} \left\| \widehat{W}_n(X_j, \check{\theta}) - W_n(X_j) \right\| \\ &\leq \frac{C_1}{2n(n-1)h^q} \sum_{i \neq j} G(Z_i) G(Z_j) K_{ij} \sup_{x \in \mathbb{R}^q} \left\| \widehat{W}_n(x, \check{\theta}) - W_n(x) \right\| \end{aligned}$$

The same argument can be applied to M_{2n} so that uniformly in $\theta \in \Theta$ and over $h, b \in \mathcal{H}_n$ and for some constant $C > 0$,

$$\left| \widehat{M}_{n,h}(\theta) - M_{n,h}(\theta) \right| \leq \frac{C}{n(n-1)h^q} \sum_{i \neq j} G(Z_i) G(Z_j) K_{ij} \sup_{x \in \mathbb{R}^q} \left\| \widehat{W}_n(x, \check{\theta}) - W_n(x) \right\|$$

The first expression on the right hand side of the last display converges in probability to $C.\mathbb{E}[G^2(Z)|X]f(X)$ which is finite by Assumption 3(ii). The result of *Step 1* then completes the proof. $\square$

7.3 Bootstrap

Lemma 7.5. *Under Assumptions 1-4, then conditionally on the sample and uniformly over $h, b \in \mathcal{H}_n$, $\sqrt{n}\big(\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h}\big)$ and $\sqrt{n}\big(\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}\big)$ have asymptotically the same distribution as $\sqrt{n}\big(\widehat{\theta}_{n,h} - \bar{\theta}_0\big)$ and $\sqrt{n}\big(\widetilde{\theta}_{n,d} - \bar{\theta}_d\big)$, respectively. That is,*

$$\sup_{h,b\in\mathcal{H}_n}\sup_{u\in\mathbb{R}}\left|\mathbb{P}(\sqrt{n}\big(\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h}\big) \leq u|\{Z_i\}_{i=1}^n) - \mathbb{P}(\sqrt{n}\big(\widehat{\theta}_{n,h} - \bar{\theta}_0\big) \leq u)\right| = o_p(1),$$
$$\sup_{b\in\mathcal{H}_n}\sup_{u\in\mathbb{R}}\left|\mathbb{P}(\sqrt{n}\big(\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}\big) \leq u|\{Z_i\}_{i=1}^n) - \mathbb{P}(\sqrt{n}\big(\widetilde{\theta}_{n,d} - \bar{\theta}_d\big) \leq u)\right| = o_p(1).$$

Proof. see section 5.2 of Lavergne & Patilea 2008 $\square$

Proof of Theorem 3

It is immediate from Lemma 7.5 that conditionally on the sample and uniformly over $h, b \in \mathcal{H}_n$, $\sqrt{n}(\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d} + \widehat{\theta}_{n,h} - \widehat{\theta}^*_{n,h})$ has asymptotically the same distribution as $\sqrt{n}(\widetilde{\theta}_{n,d} - \bar{\theta}_d + \bar{\theta}_0 - \widehat{\theta}_{n,h})$.

(i) Under H_0, we have $\bar{\theta}_d = \bar{\theta}_0 = \theta_0$ and $\widehat{Q}^*_d$ is asymptotically equivalent to $\widehat{Q}_d$ so that $HW^*_{d,h}$ and $HW_{d,h}$ have asymptotically the same $\chi^2(p)$ distribution conditionally to the sample. That is,
$\sup_{h,b\in\mathcal{H}_n}\sup_{u\in\mathbb{R}}\left|\mathbb{P}(HW^*_{d,h} \leq u|\{Z_i\}_{i=1}^n) - \mathbb{P}(HW_{d,h} \leq u)\right| = o_p(1).$

(ii) To prove the validity of the bootstrap when H_0 does not hold, consider the result given by Lemma 7.4. We note that if one replaces $g(z,\theta)$ by $wg(z,\theta)$ in all the above steps, one can easily see that the result of Lemma 7.4 also holds for the perturbed criteria $\widehat{M}^*_{n,h}(\theta)$ and $M^*_{n,h}(\theta)$. In other words, conditionally to the sample and uniformly over $h, b \in \mathcal{H}_n$ we have

$$\sup_{\theta\in\Theta}\left|\widehat{M}^*_{n,h}(\theta) - M^*_{n,h}(\theta)\right| = o_p(1). \tag{6}$$

We finally also need to show that conditionally to the sample

$$\sup_{h,b\in\mathcal{H}_n}\sup_{\theta\in\Theta}\left|M_{n,h}^*(\theta)-M_{n,h}(\theta)\right|=o_p(1) \tag{7}$$

Denote $g_n(Z,\theta)=W_n^{-1/2}(X)g(Z,\theta)$. We have

$$\begin{aligned}
h^q(M_{n,h}^*(\theta)-M_{n,h}(\theta)) &= \frac{1}{2n(n-1)}\sum_{i\neq j}(w_iw_j-1)g_n(Z_i,\theta)g_n(Z_j,\theta)K_{ij}\\
&= \frac{1}{2n(n-1)}\sum_{i\neq j}(w_i-1)(w_j-1)g_n(Z_i,\theta)g_n(Z_j,\theta)K_{ij}\\
&\quad+\frac{1}{2n(n-1)}\sum_{i\neq j}(w_i-1)g_n(Z_i,\theta)g_n(Z_j,\theta)K_{ij}\\
&\quad+\frac{1}{2n(n-1)}\sum_{i\neq j}(w_j-1)g_n(Z_i,\theta)g_n(Z_j,\theta)K_{ij}\\
&= m_{1n}(w_i,w_j)+m_{2n}(w_i)+m_{3n}(w_j)
\end{aligned}$$

Our assumptions guarantee that all the functions entering in the above terms as indexed by θ, h and b are euclidean. The term m_{1n} is a second-order degenerated U-process. It follows from Corollary 8 of Sherman (1994) that $\sup_{h,b>0}\sup_{\theta\in\Theta}|m_{1n}|=O_p(n^{-1})$. The terms m_{2n} and m_{3n} are zero-mean U-processes. By Corollary 7 of Sherman (1994), we have $\sup_{h,b>0}\sup_{\theta\in\Theta}|m_{2n}|=O_p(n^{-1/2})$ and $\sup_{h,b>0}\sup_{\theta\in\Theta}|m_{3n}|=O_p(n^{-1/2})$. Hence, $\sup_{h,b\in\mathcal{H}_n}\sup_{\theta\in\Theta}h^q\left|M_{n,h}^*(\theta)-M_{n,h}(\theta)\right|=O_p(n^{-1/2})$, so that $\sup_{\theta\in\Theta}\left|M_{n,h}^*(\theta)-M_{n,h}(\theta)\right|=o_p(1)$, uniformly over $h,b\in\mathcal{H}_n$.

It then follows from (5) (6) and (7) that

$$\sup_{h,b\in\mathcal{H}_n}\sup_{\theta\in\Theta}\left|\widehat{M}_{n,h}^*(\theta)-\widehat{M}_{n,h}(\theta)\right|=o_p(1) \tag{8}$$

We now use (8) to show that conditionally on the sample, $\widehat{\theta}_{n,h}^*-\widehat{\theta}_{n,h}=o_p(1)$ uniformly in $h,b\in\mathcal{H}_n$. By (8), we have $\widehat{M}_{n,h}^*(\widehat{\theta}_{n,h}^*)-M_{n,h}(\widehat{\theta}_{n,h}^*)=o_p(1)$ and $\widehat{M}_{n,h}^*(\widehat{\theta}_{n,h})-M_{n,h}(\widehat{\theta}_{n,h})=o_p(1)$ uniformly in $h,b\in\mathcal{H}_n$. Also, by definition,

$\widehat{M}_{n,h}(\widehat{\theta}_{n,h}) \leq \widehat{M}_{n,h}(\widehat{\theta}^*_{n,h})$ and $\widehat{M}^*_{n,h}(\widehat{\theta}^*_{n,h}) \leq \widehat{M}^*_{n,h}(\widehat{\theta}_{n,h})$. Hence,

$$\begin{aligned}
\widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) &= \widehat{M}^*_{n,h}(\widehat{\theta}^*_{n,h}) + \Big(\widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) - \widehat{M}^*_{n,h}(\widehat{\theta}^*_{n,h})\Big) \\
&= \widehat{M}^*_{n,h}(\widehat{\theta}^*_{n,h}) + o_p(1) \leq \widehat{M}^*_{n,h}(\widehat{\theta}_{n,h}) + o_p(1) \\
&= \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) + \Big(\widehat{M}^*_{n,h}(\widehat{\theta}_{n,h}) - \widehat{M}_{n,h}(\widehat{\theta}_{n,h})\Big) + o_p(1) \\
&= \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) + o_p(1) + o_p(1) = \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) + o_p(1)
\end{aligned}$$

Thus, $\widehat{M}_{n,h}(\widehat{\theta}_{n,h}) \leq \widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) \leq \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) + o_p(1)$, so that uniformly over $h, b \in \mathcal{H}_n$

$$\widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) - \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) = o_p(1) \tag{9}$$

Since $\widehat{\theta}_{n,h}$ is the minimizer of $\widehat{M}_{n,h}(\theta)$ in the compact set Θ, then we have $\forall \epsilon > 0$, $\inf_{\{\|\theta - \widehat{\theta}_{n,h}\| \geq \epsilon\}} \widehat{M}_{n,h}(\theta) > \widehat{M}_{n,h}(\widehat{\theta}_{n,h})$. In other words, $\forall \epsilon > 0$, $\exists \mu > 0$ such that $\|\theta - \widehat{\theta}_{n,h}\| \geq \epsilon$ implies $\widehat{M}_{n,h}(\theta) > \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) + \mu$. Thus, the event $\Big\{\|\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h}\| \geq \epsilon\Big\}$ is contained in the event $\Big\{\widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) - \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) > \mu\Big\}$. Since by (9) the probability of the latter converges to zero, so is the probability of the former. That is, $\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h} = o_p(1)$ uniformly in $h, b \in \mathcal{H}_n$. Likewise, all the above steps can be repeated to establish that $\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d} = o_p(1)$ uniformly in $b \in \mathcal{H}_n$ for any fixed $d > 0$. Hence,

$$\begin{aligned}
n^{-1} HW^*_{d,h} &= \Big((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h})\Big)' \widehat{Q}^{*-1}_d \Big((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h})\Big) \\
&= o_p(1),
\end{aligned}$$

and by Markov inequality, $\mathbb{P}\Big[\sup_{h,b \in \mathcal{H}_n} n^{-1} HW^*_{d,h} \geq \epsilon | Z_1, \ldots, Z_n\Big] = o_p(1)$, $\forall \epsilon$ $\square$

Table 1: **Rejection frequency of Specifications Tests for $n = 100$ at 5% level**

Models	HW^b	HW	CEL	SCEL	ZHENG	RESET
Homoskedastic						
H_0	0.052	0.069	0.059	0.041	0.004	0.010
H_1^A, $s = 0.2$	0.366	0.486	0.578	0.529	0.407	0.976
H_1^A, $s = 0.3$	0.384	0.640	0.792	0.760	0.647	0.996
H_1^A, $s = 0.4$	0.421	0.729	0.895	0.872	0.790	0.999
H_1^B, $s = 3$	0.291	0.330	0.158	0.130	0.100	0.018
H_1^B, $s = 5$	0.341	0.522	0.394	0.352	0.291	0.037
H_1^B, $s = 7$	0.430	0.566	0.703	0.676	0.547	0.042
Heteroskedastic						
H_0	0.044	0.083	0.082	0.060	0.052	0.029
H_1^A, $s = 0.2$	0.44	0.643	0.567	0.519	0.432	0.770
H_1^A, $s = 0.3$	0.76	0.893	0.769	0.740	0.641	0.94
H_1^A, $s = 0.4$	0.842	0.933	0.898	0.887	0.825	0.990
H_1^B, $s = 3$	0.109	0.228	0.390	0.344	0.268	0.074
H_1^B, $s = 5$	0.241	0.492	0.768	0.732	0.597	0.083
H_1^B, $s = 7$	0.292	0.698	0.940	0.930	0.822	0.100
Mixture						
H_0	0.047	0.079	0.077	0.053	0.058	0.033
H_1^A, $s = 0.2$	0.146	0.278	0.536	0.494	0.409	0.782
H_1^A, $s = 0.3$	0.252	0.390	0747	0.716	0.632	0.928
H_1^A, $s = 0.4$	0.312	0.419	0.874	0.854	0.789	0.980
H_1^B, $s = 3$	0.104	0.154	0.351	0.298	0.235	0.062
H_1^B, $s = 5$	0.214	0.320	0.704	0.677	0.535	0.085
H_1^B, $s = 7$	0.366	0.478	0.911	0.901	0.771	0.082

References

Antoine, B., Bonnal, H. & Renault, E. (2007), 'On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood', *Journal of Econometrics* **138**(2), 461–487.

Chamberlain, G. (1987), 'Asymptotic efficiency in estimation with conditional moment restrictions', *Journal of Econometrics* **34**(3), 305–334.

Delgado, M., Dominguez, M. & Lavergne, P. (2006), 'Consistent tests of conditional moment restrictions', *Annales d'Economie et de Statistique* **82**(1), 33–67.

Dominguez, M. & Lobato, I. (2006), 'A consistent specification test for models defined by conditional moment restrictions', *Working Paper, Universidad Carlos III De Madrid, Economic series 11* .

Donald, S., Imbens, G. & Newey, K. W. (2003), 'Empirical likelihood estimation and consistent tests with conditional moment restrictions', *Journal of Econometrics* **117**, 55–93.

Dufour, J.-M. & Khalaf, L. (2001), 'Monte carlo test methods in econometrics', *A Companion to Econometric Theory* **ed. B. Baltagi, Oxford, Blackwell Publishers**, 494–519.

Dufour, J. M. & Valery, P. (2009), 'Hypothesis tests when rank conditions fail: a smooth regularization approach', *Working Paper* .

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments', *Econometrica* **50**(4), 1029–1054.

Kitamura, Y., Tripathi, G. & Ahn, H. (2004), 'Empirical likelihood-based inference in conditional moment restriction models', *Econometrica* **72**(6), 1667–1714.

Lavergne, P. & Patilea, V. (2008), 'Smooth minimum distance estimation and testing in conditional moment restriction models: Uniform in bandwidth theory', *Working Paper* .

MacKinnon, J. (2007), 'Bootstrap hypothesis testing', *Queens Economic Department Working Paper* **1127**.

Newey, K. W. D. (1985), 'Generalized method of moments specification testing', *Journal of Econometrics* **29**, 229–256.

Nolan, D. & Pollard, D. (1987), 'U-processes: Rates and convergence', *The Annals of Statistics* **15**(2), 780 – 799.

Otsu, T. (2008), 'Conditional empirical likelihood estimation and inference for quantile regression models', *Journal of Econometrics* **142**, 508–538.

Pakes, A. & Pollard, D. (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica* **57**(5), 1027–1057.

Sherman, R. (1994a), 'Maximal inequalities for degenerate u-processes with application to optimization estimators', *The Annals of Statistics* **22**(1), 439 – 459.

Tauchen, G. (1985), 'Diagnostic testing and evaluation of maximum likelihood models', *Journal of Econometrics* **30**, 415–443.

Tripathi, G. & Kitamura, Y. (2003), 'Testing conditional moment restrictions', *The Annals of Statistics* **31**(6), 2059–2095.

Wooldridge, J. (1990), 'A unified approach to robust, regression-based specification tests', *Econometric Theory* **6**, 17–43.