WORKING PAPERS

N° 16-743

December 2016

“A Hausman Specification Test of Conditional Moment Restrictions”

Pascal Lavergne and Pierre Nguimkeu

Toulouse School of Economics

A Hausman Specification Test of Conditional Moment Restrictions*

Pascal Lavergne† Pierre Nguimkeu‡

November 2016

Abstract

This paper addresses the issue of detecting misspecified 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. An empirical application to a model of female formal labor force participation and wage determination in urban Ghana is provided.

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

JEL Classification: C52, C12, C14, C15.

*Pierre Nguimkeu gratefully acknowledges financial support from the SSHRC Canada. We thank Russell Davidson, Arthur Lewbel, Victoria Zinde-Walsh, and Karim Abadir for useful comments. We also thank conference participants of 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, the African Econometric Society, as well as seminar participants at Simon Fraser University, Georgia State University, Florida State University, Georgia Institute of Technology, and Federal Reserve Bank of Atlanta.

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

‡Corresponding author. Department of Economics, Georgia State University; PO Box 3992, Atlanta, GA 30303, USA; Phone: +1404-413-0162; Fax: +1404-413-0145; Email: nnguimkeu@gsu.edu

1 Introduction

This paper addresses the issue of detecting misspecification in 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 in terms of martingale differences. Other examples include models identified through instrumental variables, models defined by conditional mean and conditional variance without specific assumptions on their distribution, nonlinear simultaneous equation models, and transformation models. Estimation of such models have been extensively investigated. One of the most popular techniques is the generalized method of moments (GMM) introduced by Hansen (1982). But subsequent techniques have also been considered to provide more efficient and accurate estimators. Chamberlain (1987) showed that the semiparametric efficiency bound for CMR models can be attained and Robinson (1987) and Newey (1993) discussed ways to obtain this semiparametric efficiency bound using nonparametric optimal instruments. Donald, Imbens & Newey (2003), Kitamura, Tripathi & Ahn (2004), Smith (2007a) and Antoine, Bonnal & Renault (2007) proposed smoothed bandwidth-dependent Empirical Likelihood (EL) methods. Dominguez & Lobato (2004) introduced a class of estimators whose consistency does not depend on any user-chosen parameter, but cannot attain the semi-parametric efficiency bound. In a recent work, Lavergne & Patilea (2013) proposed a new class of estimators obtained by Smooth Minimum Distance (SMD) estimation, which provides an alternative to the Dominguez and Lobato's approach and allows for semiparametric efficiency. Their framework provides a way to obtain $\sqrt{n}$-consistent and asymptotically normal estimators uniformly over a wide range of bandwiths including arbitrary fixed ones, as well as a semiparametrically efficient estimator using a vanishing bandwidth.

All the above 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 drastically different properties. A central issue for the practitioner is therefore to check the validity of these restrictions. 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 almost surely. We use an approach *à la* Hausman (1978), exploiting the properties of the SMD estimators developped by Lavergne & Patilea (2013). In particular, we base our test on the distance between a consistent and asymptotically efficient SMD estimator - 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 a bootstrap method to approximate the critical values of this test in small samples. The distribution and the validity of our bootstrap procedure are studied. Simulations show that the proposed specification test has good size and power performance in small and moderate samples. We then use it in an empirical application to detect sample selection bias in a model of female labor supply in Ghana.

Other specification tests for CMR have been proposed in the literature. Some of them are based on the GMM and test a finite set of arbitrary unconditional moment restrictions implied by the conditional moment restrictions, see, e.g., Newey (1985), Tauchen (1985) and Wooldridge (1990). However, Dominguez & Lobato (2006) raised global identification issues surrounding the GMM-based tests and proposed, along with Delgado, Dominguez & Lavergne (2006) consistent specification tests based on a Cramer Von Mises criterion. But the asymptotic distribution of their test statistics depend 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. A particular advantage of our test is that it does not suffer from the possible identification issue inherent in GMM-based tests as it uses the full information contained in the definition of the model, which involves an infinite number of unconditional moments. Also, it is more versatile than most existing tests since it applies to a wide range of moment functions including non-differentiable ones as in conditional quantile regressions models. Finally, our test statistic is easy to compute since it only requires computation of a quadratic form which involves the difference of the parameter estimates and the difference of the estimated covariance matrices.

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 properties. A Bootstrap procedure to compute critical values of the test in small samples is proposed in Section 4. Section 5 reports Monte Carlo simulations results and Section 6 provides an empirical application. Section 7 concludes and Section 8 gathers all the proofs and some technical formulas.

2 Framework and Test

In this section, we describe our general framework for the specification analysis of CMR models, and we explain the rationale for the proposed test. We use the following notations throughout the paper. For a real valued function $l(\cdot)$, $\nabla_\theta l(\cdot)$ and $H_{\theta,\theta} l(\cdot)$ denote the p-column vector of the 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 a nonlinear multivariate regression. In others, $\mathbb{E}[g(Z,\theta_0)|X]$ is seen as the first order partial derivatives of a stochastic optimization problem.

Our test statistic uses the Lavergne & Patilea (2013) smooth minimum distance (SMD) class of estimators for θ_0 characterized by (1). The typical SMD estimator obtains as the argument minimizing

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

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 & Patilea (2013) 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

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

where b is a bandwidth converging to zero, which we assume to be the same as h to simplify the exposition. Likewise, we note that a different kernel could also be used in the above estimator without affecting our results, as long as this kernel satisfies the assumptions stated below.

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 involving 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), and a consistent but inefficient one, $\widetilde{\theta}_{n,d}$, that uses a fixed bandwidth d and the same weighting matrix. Hence, we define the test statictics as

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

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, $\delta_{d,h} = \widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h}$ converges in probability to zero. The test statistic then has a standard chi-squared limiting distribution. Under misspecification, the two estimators converge to different values in general, so that the distance between $\widehat{\theta}_{n,h}$ and $\widetilde{\theta}_{n,d}$ is nonzero in large samples. Hence, significantly large values of $T_{d,h}$ are regarded as evidence that the conditional moment restrictions are not consistent with the data. Thus, the α-level asymptotic test is $\mathbb{I}\left(T_{d,h} > c_\alpha\right)$ where c_α is the $1-\alpha$ quantile of a χ^2_p distribution.

A practical drawback of our test, which is typical of Hausman-type tests, is that in some instances the asymptotic variance of the estimators' differences could be singular, so that one should use a *modified* inverse, as proposed by Lutkepohl & Burda (1997), or a *regularized* inverse, as proposed by Dufour & Valery (2011). 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 one quasi-Newton step.

3 Asymptotic Distribution

We now provide regularity conditions under which the asymptotic distribution 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 matrix $\text{Var}[g(Z,\theta_0)|X=\cdot]f(\cdot)$, as shown by Lavergne & Patilea (2013). Define $\tau(x,\theta)=\mathbb{E}[g(Z,\theta)|X=x]$.

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

In particular, for any fixed $d>0$, the parameter $\bar{\theta}_d$ uniquely minimizes

$$\mathbb{E}M_{n,d}(\theta) = \frac{1}{2}\mathbb{E}\left[\tau(X_1,\theta)'W^{-1/2}(X_1)W^{-1/2}(X_2)\tau(X_2,\theta)d^{-q}K((X_1-X_2)/d)\right]$$

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 sup-*

port 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.

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 $\theta_1, \theta_2 \in \mathring{\Theta}$ 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 $\mathring{\Theta}$. There exists a real valued function $H(\cdot)$ with $\mathbb{E}H^4 < \infty$ and some constant $a \in (0,1]$ such that:
$\|H_{\theta,\theta}\tau^{(k)}(X,\theta_1) - H_{\theta,\theta}\tau^{(k)}(X,\theta_2)\| \leq H(Z)\|\theta_1 - \theta_2\|^a$, $\forall\ \theta_1, \theta_2 \in \mathring{\Theta}$, $k = 1,\ldots,r$.
(vi) The components of $\nabla_\theta\tau(\cdot,\theta_1)f(\cdot)$ and of $\mathbb{E}\left[g(Z,\theta_1)g'(Z,\theta_2)|X = \cdot\right] f(\cdot)$, $\theta_1, \theta_2 \in \mathring{\Theta}$, are uniformly bounded in $L^1 \cap L^2$ and are continuous in $\theta_1, \theta_2 \in \mathring{\Theta}$.

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 $\bar{\theta}_h = \theta_0\ \forall h$ in Assumption 1 (Lavergne & Patilea 2013). 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 for any bandwidth, including fixed ones (see Lavergne & Patilea 2013). 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 guarantees 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. Note that twice differentiability of $g(z,\cdot)$ is not needed for the construction of our Hausman test statistic. Only the differentiability of $\tau(x,\cdot)$ is needed to establish our asymptotic results. This allows the specification test to apply to a wider variety of models including, e.g., conditional quantile restrictions. Assumption 4 is needed only when studying the test's 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 & Patilea (2013). 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, $\sqrt{n}\left(\widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h}\right)$, regarded 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 $\widehat{Q}_d$, are given in Section 8. 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 $d > 0$, $T_{d,h}$ converges in distribution to a χ^2_p uniformly over $h \in \mathcal{H}_n$.*

The proposed statistic has an asymptotic chi-squared distribution under the null, so that standard statistical testing procedures can be used in large samples. Existing consistent tests such as those based on a Cramer-von-Mises criterion (Dominguez & Lobato 2006, Delgado et al. 2006) have asymptotic null distributions that depend on the underlying data generating process. This makes standard inference infeasible, and is therefore an important practical limitation. When the model is misspecified, the population conditional moment $\tau(X,\theta) = \mathbb{E}[g(Z,\theta)|X]$ is different from zero for any value of the parameter θ. In this case, the function $\mathbb{E}M_{n,0}(\theta) = \lim_{h\to 0}\mathbb{E}M_{n,h}(\theta) = \frac{1}{2}\mathbb{E}\left[\tau(X,\theta)'W^{-1}(X)\tau(X,\theta)f(X)\right]$ is not minimized at $\bar{\theta}_d$ for arbitrary values of $d > 0$, implying that $\text{Plim}_{n\to\infty}\delta_{d,h} \neq 0$ and the test statistic diverges at rate n.

4 Bootstrap Approximation

Bootstrapping is a popular approach 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 generating 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. Here, we propose a simple method that allows to circumvent these difficulties, if they appear, applies generally and is easy to implement. This method has been proposed by Jin, Ying & Wei (2001) and Bose & Chatterjee (2003), see also Chatterjee & Bose (2005) for a similar method applied to Z-estimators and Chen & D. (2009) for sieve minimum distance estimators. However, their method 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

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

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 also 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 $d > 0$ 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

$$T^*_{d,h} = n\left((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h}) \right)' \widehat{Q}^{*-1}_d \left((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h}) \right) ,$$

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 $\{T^*_{d,h,j}\}_{j=1}^B$. This bootstrap empirical distribution is then used to approximate the distribution of the test statistic $T_{d,h}$ under correct specification, allowing one to calculate the critical values empirically. Typically, one rejects H_0 at α level if $T_{d,h} > c^*_{\alpha,B}$, where $c^*_{\alpha,B}$ is the upper α-percentile of the empirical distribution $\{T^*_{d,h,j}\}_{j=1}^B$.

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 increases do not substantially affect the critical values. MacKinnon (2009) pointed out that the number of bootstrap samples B must also be such that the quantity $\alpha(B+1)$ is an integer, where α is the level of the test. Moreover, as explained 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 2. *Let Assumptions 1-3 hold. Then, conditionally on the sample,*
(i) $\sup_{h\in\mathcal{H}_n}\sup_{u\in\mathbb{R}}\left|\mathbb{P}(T^*_{d,h}\leq u|\{Z_i\}_{i=1}^n)-\mathbb{P}(T_{d,h}\leq u)\right| = o_p(1)$*, under* H_0*;*
(ii) $T^*_{d,h} = o_p(n)$ *uniformly over* $h\in\mathcal{H}_n$*, when* H_0 *does not hold.*

The first part of the theorem implies that the level α critical value $c^*_{\alpha,B}$ obtained from the bootstrap distribution of $T^*_{d,h}$ converges (conditional on the original sample) to the critical value c_α from the limiting distribution of $T_{d,h}$ as $B\to\infty$ and $n\to\infty$. This suggests that the asymptotic significance level of our test using the bootstrap critical values is as desired. Since $T_{d,h}$ diverges at rate n under the alternative, the second part of the theorem implies that $\mathbb{P}[T_{d,h} > T^*_{d,h}]\xrightarrow{p}1$ when $n\to\infty$, which suffices for consistency.

5 Monte Carlo Simulations

In this section we conduct Monte Carlo simulations to provide evidence on the behavior of our test statistic in small samples, and compare our results with some existing tests. Two simulations are performed, one with a regression model, the other with a binary choice model.

5.1 Simulation Study 1

The set up of this simulation is a regression model. Our main focus is to examine the behavior of the specification test statistic under the null that the model is correctly specified, then assess its properties under a set of alternative hypotheses. Throughout this simulation, the null hypothesis is:

$$H_0: \quad \mathbb{E}[Y - \theta_1 - \theta_2 X | X] = 0 \quad \text{a.s. for some } (\theta_1, \theta_2) \tag{5}$$

where X and Y are univariate random variables. The variables are randomly generated from the following data generating processes:

$$Y = \theta_1 + \theta_2 X + \nu, \tag{6}$$

and

$$Y = \theta_1 + \theta_2 X + s\boldsymbol{\lambda}\left((\theta_1 + \theta_2 X)/s\right) + \nu, \quad s = 0.3, 0.5, 0.7 \tag{7}$$

where $\boldsymbol{\lambda}(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the Inverse Mill's ratio. In this formula, $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability and cumulative density functions respectively. The parameters are set to $\theta_1 = \theta_2/2 = 1$ and $X \sim N(0, 1)$. As for the error term ν, we consider two different situations:

- Homoskedastic errors, i.e. $\nu = \epsilon$, where $\epsilon \sim N(0, 1)$ and ϵ is independent of X.
- Heteroskedastic errors, i.e. $\nu = \epsilon\sqrt{.1 + .1X^2}$,

When the data is generated from Equation 6, the model being tested is correctly specified. When it is generated from Equation 7, the model is misspecified so that the hypothesis H_0 is false and $\mathbb{E}[Y|X]$ has both a linear term and a nonlinear term given by the Inverse Mill's ratio. Different values of s correspond to different degrees of deviation from the null. These alternatives mimic situations where the regression model may be suffering from a specification error in the sense of Heckman (1979), perhaps due to sample selection (see, e.g. Greene 2012, pp. 837-839).

Our specification test statistics are computed using a fixed bandwidth $d = 1$ for the consistent estimator, while the size-dependent bandwidth used for the efficient estimator is taken as $h = cn^{-1/5}$, for some constant $c > 0$. The sensitivity to this constant is assessed by considering different values, $c = 0.8; 1.0; 1.3; 1.5$.[1] The gaussian kernel is used for both the asymptotic test and the bootstrap approximation. The empirical sizes and powers of the tests are computed at both the 5% and the 10% nominal levels with sample sizes $n = 100$ and $n = 50$, using $1,000$ replications.

[1] We only present results for $c = 1.5$ and $c = 1$; results for $c = 0.8$ and $c = 1.3$ are available upon request.

Table 1: Rejection frequency of the Proposed Tests

	$T^*_{d,h}$		$T_{d,h}$		$T^*_{d,h}$		$T_{d,h}$	
Models	$c = 1.5$	$c = 1$	$c = 1.5$	$c = 1$	$c = 1.5$	$c = 1$	$c = 1.5$	$c = 1$
$\alpha = 5\%$			$n = 100$				$n = 50$	
Homoskedastic								
H_0	0.007	0.003	0.025	0.016	0.004	0.007	0.049	0.027
H_1, $s = 0.3$	0.601	0.486	0.953	0.823	0.177	0.138	0.708	0.522
H_1, $s = 0.5$	0.584	0.475	0.924	0.794	0.156	0.112	0.694	0.507
H_1, $s = 0.7$	0.580	0.473	0.927	0.811	0.152	0.110	0.633	0.476
Heteroskedastic								
H_0	0.027	0.016	0.081	0.043	0.009	0.012	0.082	0.088
H_1, $s = 0.3$	0.950	0.925	1.000	0.990	0.537	0.430	0.980	0.936
H_1, $s = 0.5$	0.956	0.929	0.999	0.989	0.527	0.427	0.977	0.931
H_1, $s = 0.7$	0.945	0.945	0.999	0.993	0.535	0.420	0.977	0.927
$\alpha = 10\%$			$n = 100$				$n = 50$	
Homoskedastic								
H_0	0.026	0.026	0.076	0.061	0.036	0.027	0.076	0.071
H_1, $s = 0.3$	0.818	0.695	0.979	0.899	0.382	0.338	0.777	0.648
H_1, $s = 0.5$	0.803	0.686	0.962	0.869	0.365	0.332	0.770	0.634
H_1, $s = 0.7$	0.768	0.696	0.971	0.884	0.372	0.318	0.738	0.604
Heteroskedastic								
H_0	0.076	0.053	0.132	0.106	0.061	0.049	0.138	0.154
H_1, $s = 0.3$	0.989	0.965	1.000	0.992	0.758	0.666	0.987	0.962
H_1, $s = 0.5$	0.986	0.966	0.999	0.994	0.762	0.654	0.988	0.951
H_1, $s = 0.7$	0.989	0.976	0.999	0.994	0.751	0.678	0.988	0.950

Table 1 summarizes both our general test statistic and bootstrap results. The figures reported on the table are simulated rejection probabilities. The first row of each model reports simulation results under the null, H_0 - thus showing the empirical size of each test - and the remaining rows report simulation results under various alternatives, H_1. Our general test, denoted $T_{d,h}$, displays both a reasonable size and a remarkably good power under the various alternatives.

For our bootstrap test, denoted $T^*_{d,h}$, we compute 199 bootstrap statistics from 1000 replications with the sample sizes of $n = 50$ and $n = 100$. At each replication, critical values at 5% (respectively, 10%) significance are calculated by taking the 95^{th} (respectively, 90^{th}) upper percentiles 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:

$$\mathbb{P}\left[\omega_i = \frac{3-\sqrt{5}}{2}\right] = 1 - \mathbb{P}\left[\omega_i = \frac{3+\sqrt{5}}{2}\right] = \frac{5+\sqrt{5}}{10} \tag{8}$$

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 good bootstrap approximations of the test statistic. As reported in Table 1 our bootstrap test has good empirical sizes since all rejection probabilities are within the nominal range of 5% and 10% accordingly. Note, however, that the bootstrap test appears to be somewhat conservative, although this does not deteriorate the power. This implies that the practitioners need not fear that the rate of type I errors would exceed the nominal rate and lead to invalid conclusion about the model specification. The power performance of the bootstrap test is also fairly good, though worse than our asymptotic test. This feature is however expected since a gain in size is often traded off with a relative loss in power in the bootstrap test due to its conservative nature. The size of the bootstrap test turns out to be very robust to various values of the constant $c > 0$ while the values of the asymptotic test are more sensitive to this constant especially for the smaller sample size of $n = 50$. A possible explanation is that while the critical values of the asymptotic test are fixed upfront, the critical values of the bootstrap test incorporate the variations in bandwidths.

5.2 Simulation Study 2

In this simulation study, we examine the performance of our test for binary choice models and compare it with some existing tests such as the Horowitz & Härdle (1994) test statistic (denoted HH) and the Härdle, Mammen & Proença (2001) bootstrap test statistics (denoted HMP). Binary choice models are examples of single index models where the response variable, Y, takes on two possible values, 0 and 1. They are defined by

$$\Pr[Y = 1|X] = v(X'\theta), \quad \text{where } v(\cdot) \text{ is the link function.}$$

These models can be rewritten in the form of conditional moment restrictions as $\mathbb{E}[g(Z,\theta)|X] = 0$, where $Z = (Y, X')'$ and $g(Z,\theta) = Y - v(X'\theta)$.

We follow the simulation set-up of Härdle et al. (2001). The null hypothesis is defined by

$$H_0 : \mathbb{E}[Y|X] = \{1 + \exp(-1 - \theta_1 X_1 - \theta_2 X_2)\}^{-1}$$

where X_1 and X_2 are independent standard normal random variables and the parameter values are $\theta_2 = 2\theta_1 = 2$. The null hypothesis therefore assumes that the data come from a logit model. The alternative hypothesis considered is a family of link functions called *logit with bump*, defined by

$$H_1 : \mathbb{E}[Y|X] = \{1 + \exp(-1 - \theta_1 X_1 - \theta_2 X_2)\}^{-1} - \frac{a}{1.5}\phi\left(\frac{a}{1.5}\right)$$

where $\phi(\cdot)$ is, as before, the pdf of the standard normal and $a \in \{0.75; 1; 1.25\}$. Two sample sizes are considered, $n = 200$ and $n = 500$, along with 199 bootstraps and 500 replications. For the proposed specification test, we use the gaussian kernel for both the asymptotic and the bootstrap approximation. The bandwidth for the consistent estimator is fixed at $d = 1$ while the vanishing bandwidth is taken as in Härdle et al. (2001) at $h = cn^{-1/5}$, with $c = 1.44$ and $c = 4.33$.[2] The bootstraps samples are generated as in the previous simulation example using the two-points distribution given by Equation 8. The nominal size of the tests are fixed at 10% and 5%. For the HH and HMP tests, we only present the one-sided tests which have overall much better performance than their two-sided counterparts.

Table 2 reports the asymptotic and the bootstrap rejection probabilities of the proposed asymptotic test and bootstrap ($T_{d,h}$ and $T^*_{d,h}$) as well as the Horowitz & Härdle, and Härdle et al. tests (HH and HMP) at both the 10% and 5% significance levels for a sample size of 200 observations. The empirical size of the tests are given in the rows corresponding to $H_0, a = 0.00$ while the empirical powers are given in the remaining rows corresponding to H_1 at different levels of deviations from the null, $a = 0.75$, $a = 1.00$, and $a = 1.25$. The proposed tests have sizes that are comparable to the HH test but are overall smaller than the size of the HMP test. As for the empirical powers, the HMP test displays powers that are fairly acceptable, while the powers of the HH test are clearly unsatisfactory. In fact, the HH test is unable to reject the misspecified models at both the 5% and 10% level. In contrast, both

[2] Härdle et al. (2001) also perform their test with $c = 2.88$. But this corresponds to $h = 1$, our fixed bandwidth.

Table 2: Rejection frequency of specification tests for $n = 200$

Models	$T^*_{d,h}$		$T_{d,h}$		HMP		HH	
	$c = 1.4$	$c = 4.3$	$c = 1.4$	$c = 4.3$	$c = 1.4$	$c = 4.3$	$c = 1.4$	$c = 4.3$
Size $\alpha = 10\%$								
H_0, $a = 0.00$	0.002	0.001	0.006	0.020	0.098	0.146	0.022	0.000
H_1, $a = 0.75$	0.630	0.380	0.930	0.996	0.150	0.390	0.024	0.012
H_1, $a = 1.00$	0.704	0.756	0.954	1.000	0.266	0.488	0.048	0.030
H_1, $a = 1.25$	0.766	0.658	0.952	1.000	0.416	0.608	0.106	0.064
Size $\alpha = 5\%$								
H_0, $a = 0.00$	0.001	0.000	0.006	0.016	0.048	0.074	0.016	0.000
H_1, $a = 0.75$	0.172	0.142	0.7920	0.996	0.150	0.390	0.024	0.000
H_1, $a = 1.00$	0.408	0.492	0.854	0.996	0.168	0.344	0.026	0.008
H_1, $a = 1.25$	0.350	0.372	0.878	1.000	0.286	0.424	0.068	0.022

the asymptotic and the bootstrap versions of the proposed test have remarkably better power, and significantly outperform both the HH and the HMP tests. These results are consistent regardless of the constant on the vanishing bandwidth, as shown in the theory. This suggests that our tests may have a stronger ability to detect misspecification in these types of models than the above competitors. When a larger sample size of 500 observations is used, the simulations results show that our tests perform even better, with the powers of the asymptotic tests consistently hitting the limit of 1 (see Table 3).

Table 3: Rejection frequency of specification tests for $n = 500$ and $c = 2.8$

Models	$T^*_{d,h}$	$T_{d,h}$	HMP	HH
Size $\alpha = 10\%$				
H_0, $a = 0.00$	0.001	0.014	0.114	0.010
H_1, $a = 0.75$	0.610	1.000	0.538	0.106
H_1, $a = 1.00$	0.966	1.000	0.790	0.268
H_1, $a = 1.25$	0.968	1.000	0.908	0.568
Size $\alpha = 5\%$				
H_0, $a = 0.00$	0.001	0.004	0.062	0.010
H_1, $a = 0.75$	0.712	1.000	0.396	0.062
H_1, $a = 1.00$	0.996	1.000	0.658	0.196
H_1, $a = 1.25$	0.980	1.000	0.842	0.418

To sum up, our general test statistic has very good power performance in our simulation experiments and are competitive with existing tests. Moreover, the empirical size performance of our tests shows that our bootstrap test can properly handle small sample size models.

6 Empirical Application

To illustrate the use of the test statistic developed in this paper, we apply it to a model of female labor supply and wage determination in urban Ghana. The model is a variant of the classic model of Mroz (1987) and is applied to a cross-sectional random representative sample of 1804 Ghanaian women, 277 of which participated in the formal labor market. The data set contains information about wages, hours of work, experience, education, and demographic characteristics such as age, number of children, other income, etc. All data are taken from the Ghanaian Standard Living Survey 2005-2006 available at the World Bank website at http://microdata.worldbank.org/index.php/catalog/1064.

We first consider the subsample of working women and specify a simple wage equation:

$$\log Wage_i = \beta_1 + \beta_2 Educ_i + \beta_3 Exper_i + \beta_4 Exper_i^2 + \varepsilon_i$$

where $Educ_i$ is education, $Exper_i$ is labor market experience of the ith woman, and $\varepsilon_i \sim iid(0, \sigma^2)$. We estimate and test this model without controlling for selection into the formal labor market. The results of the OLS and SMD estimation as well as the value of our general specification test are presented in the first panel of Table 4. In computing our test, the fixed bandwidth was taken at $d = 1$ (the corresponding SMD estimator is denoted SMD_d), the size-dependent bandwidth taken at $h = n^{-1/5}$ (the corresponding SMD estimator is denoted SMD_h) and the gaussian kernel was used throughout. The p-value of the specification test $T_{d,h}$ is estimated at 0.0072, suggesting that the model is misspecified at the 1% significance level. The more obvious source of misspecification is the fact that the estimates obtained in the above model are constructed from a sample of working women without accounting for self-selection into the formal labor force. This sample contains only 15.4% of the women of our whole sample and, as is well known, is likely to yield estimates that are inconsistent due to the correlation between the regressors and the error induced by the sample selection mechanism.

To further assess the power of our specification test in this application, we specify a selection-corrected version of the above wage equation as suggested by Heckman (1979), and estimate and test this equation with our specification test statistic. Since the latter model is an improvement over the former, so should be the p-value of the specification test. Following Heckman (1979), we can account for sample selection bias by associating to the above equation a participation equation as follows. Consider the latent variable U representing, say, the desire to participate in the formal labor market or the difference between the offered labor market wage and the reservation wage. The ith woman participates only if $U_i > 0$. The selection is therefore defined by the dummy variable $D_i = \mathbb{I}(U_i > 0)$ where the labor force participation equation is defined by

$$U_i = \gamma_1 + \gamma_2 Educ_i + \gamma_3 Exper_i + \gamma_4 Exper_i^2 + \gamma_5 Kids_i + \gamma_6 Otherinc_i + \nu_i.$$

Here, $Kids$ is the number of young children, $Otherinc$ is the log of other income including husband's income and/or remittances from family and relatives. These variables can be seen as exclusion restrictions that influence participation but not hourly wages. As usual, we assume $\nu_i \sim N(0,1)$. Denote $\mathbf{x} = [1, Educ, Exper, Exper^2]'$, $\mathbf{w} = [1, Educ, Exper, Exper^2, Kids, Otherinc]'$, $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]'$, and finally $\gamma = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6]'$. Using both the wage equation and the participation equation above, the selection-corrected model for the working women can be rewritten (see, eg. Greene 2012, pp. 873-876) as

$$\mathbb{E}\left[\log Wage_i - \mathbf{x}_i'\beta - \rho\sigma\boldsymbol{\lambda}(\mathbf{w}_i'\gamma)\middle|\mathbf{x}_i, \mathbf{w}_i\right] = 0$$

where ρ is the correlation between ε_i and ν_i and $\boldsymbol{\lambda}(\cdot)$ is the Inverse Mill's ratio. Smooth minimum distance and Heckman two-step estimates are reported in the second panel of Table 4. Only the parameters of the wage equation are shown in the table. Note that for the two-step method, the estimates of the joint coefficient $\rho\sigma$ is first obtained and the underlying structural individual parameters ρ and σ are then deduced by the method of moments. The SMD estimation computes the estimates of these structural parameters directly. As expected, the differences between the SMD estimates are larger in the initial wage equation than in the selection-corrected equation, as are the differences between the naive OLS estimates and the Heckman's two-step estimates.

Table 4: Estimated Initial and Selection-Corrected Wage Equations*

Parameters	Initial Wage Equation			Corrected Wage Equation		
	$Least - Squares$	SMD_d	SMD_h	$Two - Step$	$SMDd$	$SMDh$
β_1	4.7520	9.8614	10.1362	-77.058	-68.2782	-78.5657
	(1.0031)	(0.001)	(0.3450)	(25.865)	(10.4055)	(11.7241)
β_2	0.2961	0.2400	0.2241	3.0339	2.7809	3.1025
	(0.0365)	(0.0057)	(0.0097)	(0.8657)	(0.3410)	(0.3823)
β_3	0.2778	-0.0269	-0.0312	0.3271	0.1469	0.1845
	(0.0821)	(0.0223)	(0.0260)	(0.0822)	(0.0409)	(0.0448)
β_4	-0.0027	0.0021	0.0020	-0.0048	-0.0025	-0.0033
	(0.0017)	(0.0004)	(0.0005)	(0.0018)	(0.0009)	(0.0010)
$(\rho\sigma)$				39.697	37.5084	42.4518
				(12.542)	(4.9547)	(5.6042)
ρ				0.799	0.7912	0.7892
					(0.0109)	(0.1561)
σ	3.3108	3.5201	3.5360	39.739	47.4070	42.9153
					(5.0123)	(4.0172)
p-Values	—	$pT_{d,h}$= 0.0072		$p_{\rho\sigma} = 7.8 \times 10^{-4}$	$pT_{d,h}$= 0.0115	

*Standard errors are in parenthesis

The selection-corrected model is tested using both our specification test and a t-test of $\rho\sigma$, the covariance of ε_i and ν_i. The p-value of this t-test is 0.0008, therefore rejecting the hypothesis that $\rho\sigma$ is zero. This confirms the presence of a selection bias in the initial wage equation estimates as initially suggested by the p-value of our specification test for the wage equation obtained earlier. Moreover, the p-value of our specification test statistic for the selection-corrected model is now estimated at 0.0115, a much higher value than the one obtained for the initial wage equation. This implies that the new model is a significant improvement over the initial one, and shows that we do not have enough evidence to reject the selection-corrected Ghanaian female wage determination equation at the 1% significance level.

7 Conclusion

This paper provides 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 (2013) Smooth Minimum Distance estimators for Conditional Moment Restrictions. The test statistic is asymptotically chi-squared under the null hypothesis 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 it to compute critical values of our test. Both the test statistic and its bootstrap counterpart are simple to implement and two Monte Carlo simulations studies are provided to show that they perform well in small and moderate samples. Moreover, the test is versatile and applies to a wide range of estimating functions including non-differentiable ones. An empirical application to a model of female formal labor force participation and wage determination in Ghana is provided to illustrate the practical usefulness of our test.

8 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.

8.1 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 8.1. *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) $\sqrt{n}\big(\widetilde{\theta}_{n,d} - \widehat{\theta}_{n,h}\big)$ *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 (2013). 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

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

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

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

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 8.2. *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 8.3. *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 (2013).
For part (ii), apply Lemma 8.2 with $A = \mathbb{E}\left[W^{-1/2}(X_2)\nabla'_\theta\tau(X_2,\theta_0)d^{-q}K((X - X_2)/d)\right] f^{-1/2}(X)$ and $B = W^{-1/2}(X)\nabla'_\theta\tau(X,\theta_0)f^{1/2}(X)$. The desired conclusion then follows. □

8.2 Asymptotic behavior of the test

Proof of Theorem 1

The result follows from Lemmas 8.1 and 8.3.□

Lemma 8.4. *Let Assumptions 1-3 hold. Then uniformly over $h, b \in \mathcal{H}_n$,*

$$\sup_{\theta\in\Theta}\left|\widehat{M}_{n,h}(\theta) - M_{n,h}(\theta)\right| = o_p(1), \tag{9}$$

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 $\bar{\theta}$,

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

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 (9). 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. □

8.3 Bootstrap

Lemma 8.5. *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 2013 □

Proof of Theorem 2

It is immediate from Lemma 8.5 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 $T^*_{d,h}$ and $T_{d,h}$ have asymptotically the same $\chi^2(p)$ distribution conditional on the sample and uniformly over h and b. That is,
$\sup_{h,b \in \mathcal{H}_n} \sup_{u \in \mathbb{R}} \left|\mathbb{P}(T^*_{d,h} \leq u|\{Z_i\}_{i=1}^n) - \mathbb{P}(T_{d,h} \leq u)\right| = o_p(1)$.

(ii) To prove the validity of the bootstrap when H_0 does not hold, consider the result given by Lemma 8.4. 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 8.4 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

$$\sup_{\theta \in \Theta} \left|\widehat{M}^*_{n,h}(\theta) - M^*_{n,h}(\theta)\right| = o_p(1). \tag{10}$$

We finally also need to show that conditionally to the sample

$$\sup_{h,b\in\mathcal{H}_n}\sup_{\theta\in\Theta}\left|M^*_{n,h}(\theta)-M_{n,h}(\theta)\right| = o_p(1) \tag{11}$$

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 (9) (10) and (11) that

$$\sup_{h,b\in\mathcal{H}_n}\sup_{\theta\in\Theta}\left|\widehat{M}^*_{n,h}(\theta)-\widehat{M}_{n,h}(\theta)\right| = o_p(1) \tag{12}$$

We now use (12) 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 (12), 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,

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

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 $h, b \in \mathcal{H}_n$

$$\widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) - \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) = o_p(1) \tag{13}$$

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 $\left\{\|\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h}\| \geq \epsilon\right\}$ is contained in the event $\left\{\widehat{M}_{n,h}(\widehat{\theta}^*_{n,h}) - \widehat{M}_{n,h}(\widehat{\theta}_{n,h}) > \mu\right\}$. Since by (13) 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,

$$\begin{aligned} n^{-1}T^*_{dh} &= \left((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h})\right)' \widehat{Q}^{*-1}_d \left((\widetilde{\theta}^*_{n,d} - \widetilde{\theta}_{n,d}) - (\widehat{\theta}^*_{n,h} - \widehat{\theta}_{n,h})\right) \\ &= o_p(1), \end{aligned}$$

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

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.

Bose, A. & Chatterjee, S. (2003), 'Generalized bootstrap for estimators of minimizers of convex functions', *Journal of Statistical Planning and Inference* **117**(2), 225–239.

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

Chatterjee, S. & Bose, A. (2005), 'Generalized bootstrap for estimating equations', *Annals of Statistics* **33**(1), 414–436.

Chen, X. & D., P. (2009), 'Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals', *Journal of Econometrics* **152**(1), 46–60.

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. (2004), 'Consistent estimation of models defined by conditional moment restrictions', *Econometrica* **72**(5), 1601–1615.

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, *in* B. Baltagi, ed., 'A Companion to Econometric Theory', Oxford, Blackwell Publishers, pp. 494–519.

Dufour, J. M. & Valery, P. (2011), 'Hypothesis tests when rank conditions fail: a smooth regularization approach', *Working Paper* .

Greene, W. (2012), *Econometric Analysis*, Prentice Hall.

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

Härdle, W., Mammen, E. & Proença, I. (2001), 'A boostrap test for single index models', *Statistics* (35), 427–451.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.

Heckman, J. (1979), 'Sample selection as a specification error', *Econometrica* **47**(1), 153–161.

Higham, N. (2008), *Functions of Matrices*, SIAM.

Horowitz, J. L. & Härdle, W. (1994), 'Testing a parametric model against a semiparametric alternative', *Econometric Theory* (10), 821–848.

Jin, Z., Ying, Z. & Wei, L. (2001), 'A simple resampling method by perturing the minimand', *Biometrika* **88**(2), 381–390.

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. (2013), 'Smooth minimum distance estimation and testing in conditional moment restriction models: Uniform in bandwidth theory', *Journal of Econometrics* **177**, 47–59.

Lutkepohl, H. & Burda, M. M. (1997), 'Modified wald tests under nonregular conditions', *Journal of Econometrics* **78**, 315–332.

MacKinnon, J. (2009), 'Bootstrap hypothesis testing', *Handbook of Computational Econometrics* .

Mammen, E. (1992), 'When does bootstrap works? asymptotic results and simulations', *Lecture notes in Statistics, Springer Verlag: New York* .

Mroz, T. (1987), 'The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions', *Econometrica* **55**(4), 765–799.

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

Newey, K. D. (1993), Efficient estimation of models with conditional moment restrictions., *in* C. R. G.S. Maddala & H. Vinod, eds, 'Handbook of Statistics', Vol. 11, pp. 419–454.

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.

Robinson, P. (1987), 'Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form', *Econometrica* **55**(4), 875–891.

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

Smith, R. (2007a), 'Efficient information theoretic inference for conditional moment restrictions', *Journal of Econometrics* **138**(2), 430–460.

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.