The University of Manchester

Economics Discussion Paper Series EDP-0912

On the Sensitivity of Kernel-based Tests of Conditional Moment Restrictions

Eduardo Fé-Rodríguez
Chris D. Orme

July 2009

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

On the Sensitivity of Kernel-based Tests of Conditional Moment Restrictions*

Eduardo Fé-Rodríguez†
Health Economics Research at Manchester

Chris D. Orme‡
Economics, School of Social Sciences

University of Manchester
Manchester, M13 9PL
United Kingdom

July 17, 2009

Abstract

This paper extends the analyses of Godfrey and Orme (1996), on the behaviour of parametric conditional moment tests in the presence unconsidered local alternatives, to that of kernel-based tests of conditional moment restrictions. Particular attention is paid to tests of separate moment conditions. The theoretical results presented in this paper provide a method of identifying whether, or not, a given test will be sensitive to a particular type of (unconsidered) local model misspecification and how any such sensitivity manifests itself. Insensitivity helps isolate possible sources of misspecification, whilst a characterisation of the sensitivity can aid the construction of robust procedures.

1 Introduction

This paper is concerned with testing the adequacy of a specified parametric model (also called the null model), with the motivation being to provide an analysis of a particular class of (non-parametric based) statistical procedures which can be used for this purpose. The class of tests considered is, itself, a subset of a much larger collection of procedures that have been proposed as "consistent" tests of (particular) Conditional Moment (hereafter CM) restrictions. They are described as consistent because the technology employed ensures that each test procedure will be consistent against all model misspecifications which *induce* failure in any of the CM restriction(s) that they are designed test. For example, a consistent test of a specified regression functional form is consistent against any failure of that specification. Thus, these procedures can equivalently be called consistent tests *for the failure of* particular CM restriction(s). The CM restrictions under scrutiny can be, either, explicit in the estimation of parametric models (e.g., the conditional zero mean

*This is a major revision of a previously circulated paper enititled "On the sensitivity of Kernel-based Conditional Moment Tests to Unconsidered Local Alternatives", and develops part of Eduardo Fé's Ph.D. Dissertation at the University of Manchester. He thanks Javier Hidalgo and Len Gill for helpful comments

†email: e.fe-rodriguez@manchester.ac.uk

‡Corresponding author; email: chris.orme@manchester.ac.uk

assumption for the error in a regression model) or implied by the parametric model (but not directly exploited in the estimation of key parameters).

Following the influential papers by Bierens (1982, 1990), two approaches have been taken (broadly speaking): those which employ smooth kernel-based based methods, or those which employ integral transforms. For example, and to name but a few, Eubank and Spiegelman (1990), Härdle and Mammen (1993), Fan and Li (1996), Zheng (1996), Ellison and Ellison (2000) and Hsiao and Li (2001), adopt the former, whilst Bierens and Ploberger (1997), Stute (1997), Stinchcombe and White (1998), Whang (2000, 2001), adopt the latter. Initially, and for the most part, the particular focus has been regression functional form through testing the assumed parametric mean specification of the dependent variable of interest. However, Hong (1993), Zheng (1994) and Hsiao and Li (2001) describe consistent tests for conditional heteroskedasticity in parametric regression models, which assume correct regression functional form. Stinchcombe and White (1998) consider a consistent test of the information matrix equality, conditional on regressors, in a fully specified parametric model and Whang (2001) develops tests of more general CM restrictions in order to assess the validity of parametric assumptions. These can be viewed as examples of testing CM restrictions that are implied by the specified null model which, as such, might be exploited for post-estimation inferential purposes (e.g., in the calculation of parameter estimate standard errors). More recently, Delgado, Dominguez and Lavergne (2006) develop consistent tests (both smooth kernel and integral transform based) of the multiple CM restrictions that define a specified parametric model; their framework encompasses all extremum estimation methods, such as generalised method of moments and quasi maximum likelihood.

The above, and related literature, provides applied workers with, potentially, an array of what we shall term Consistent Conditional Moment (hereafter CCM) tests (joint or separate) which might be employed, post estimation, to assess the validity of CM restrictions (simultaneously or separately) explicit in, or implied by, a specified parametric model. However, the analyses to date (and to the best of our knowledge) provide no general theoretical framework for describing the sensitivity of separate CCM tests to the failure of other CM restrictions that they are not specifically designed to test. For example, if the joint test proposed by Delgado et al (2006) is disaggregated into a procedure which assesses the significance of separate tests of the individual CM restrictions employed in estimation (as they suggest in their Example 1), then the effect of one source CM restriction failure on the behaviour of each separate CCM test has not been identified. In addition, when assessing the validity of CM restrictions not employed in estimation the effect of general model misspecification on the asymptotic behaviour of the appropriate CCM test has not been analysed.

From a practical point of view, results about such sensitivity (or insensitivity) could be of use if a researcher wishes to employ a CCM test to identify and characterise the failure of particular CM restrictions; e.g., correct regression functional form, conditional homoskedasticity and/or conditional symmetry. In such a situation, and as a complement to the consistency property of the CCM test (against its own particular CM restriction failure), it would be desirable, and useful, to know whether the test remains insensitive to alternative model misspecifications of the model (that it was not designed to test) and, if not, how it might (potentially) be made robust.[1] Therefore, evidence concerning the behaviour of these tests under different sources of misspecification is important. As an example, a particular point of departure is the result in Godfrey and Orme (1996, Example 4.3) which demonstrates that, in the general linear model, a parametric test for heteroskedasticity (e.g., Koenker's, 1981, test) is insensitive to local (regression) functional form misspecification (or omitted variables): is there similar insensitivity for CCM tests of conditional homoskedasticity?

With these issues in mind, this paper employs asymptotic local analysis in order to provide some qualitative guidance as to the sensitivity of CM restriction tests to *unconsidered local alternatives.*[2] Attention is restricted to a class of smooth kernel-based tests, although the "asymptotic local" methodology could also be used to analyse tests based on integral transforms. As in Godfrey and Orme (1996), and because the specified model is parametric, the approach taken is to allow misspecification by entertaining (hypothetical) alternative parametric models (i.e., sources of misspecification) to characterise the true (but unknown) data generation process which: (i) are local generalisations of the specified parametric model, and (ii) could imply violations of all, some or none of the CM restrictions under test. The more general parametric model is simply a convenient artificial device, or analytical tool, which can be used to investigate the effect of different types of misspecification and is consistent with existing Monte Carlo studies in the literature, all of which provide evidence on the power of these CCM test procedures by considering generalisations of the estimated parametric model; see, for example, Zheng (1996), Whang (2000), Ellison and Ellison (2000), Hsiao and Li (2001) and Delgado et al (2006), amongst many others. The resulting asymptotic local analysis is new since (i) it provides an analytical framework in which to place these Monte Carlo studies; (ii) it sheds light on the effect of different sources of misspecification; and, (iii) it requires a treatment of the (potentially) differing rates of convergence of the parameter estimator, when the specified null model is correct and under the entertained local alternatives, in order to identify sources of asymptotic local sensitivity. In particular, existing analyses are silent, regarding (ii), because only local violations of the specific CM restrictions under test have been considered and not sources of local model misspecification which might lead to such violations.

Intuition suggests that CCM tests should be sensitive to all misspecifications which imply a failure of any of the particular CM restrictions under test. Indeed, under various scenarios described by Godfrey and Orme (1996), where certain parametric tests are shown to be insensitive to local alternatives, it is found the "corresponding" CCM tests are sensitive; *but not in all cases.* Moreover the analysis suggests how a CCM test might be made robust to (unconsidered) local alternatives. Section 2 contains the main result of the paper. Some implications are discussed in Section 3, and some Monte Carlo evidence on the efficacy of the theoretical predictions is presented in Sections 4 and 5. Section 6 concludes.

2 The Asymptotic Behaviour of the Test Statistic

In order not to obfuscate the substantive conclusions, independently and identically distributed data and continuous conditioning/regressor variables are assumed (as in Delgado et al (2006)) and the ensuing analysis extends existing studies of local power of kernel-based test statistics (e.g., Zheng, 1996). Thus, let $\{W_i\}_{i=1}^{n}$ be a simple random sample drawn on a random variable W from an unknown Data Generation Process (DGP), and let the continuous random variable $X \in \mathbb{R}^k$ be a subvector of W with probability density $f(.)$. The asymptotic behaviour of any test criterion, constructed under some null hypothesis, will depend upon the true (but, in general) unknown DGP.

2.1 The True DGP & Assumptions

The unknown DGP is indexed by a $(p \times 1)$ parameter vector partitioned as $\varphi = (\theta', \gamma')' \in \Theta \times \Gamma \subset \mathbb{R}^p$, with true value $\varphi_0 = (\theta_0', \gamma_0')'$, and is defined by a parametric model in terms of the estimation criterion $Q_n(\varphi)$, which yields $\tilde{\varphi} \equiv \arg\max_{\varphi} Q_n(\varphi)$. Throughout, unless stated otherwise, expectations are taken with respect the true DGP. It is assumed that standard regularity conditions support the following assumptions:

Assumption A

1. $\tilde{\varphi} - \varphi_0 \xrightarrow{p} 0$, and φ_0 lies in the interior of the compact and convex parameter space $\Theta \times \Gamma \subset \mathbb{R}^p$.

2. $\sqrt{n}\partial Q_n(\varphi_0)/\partial\varphi = O_p(1)$.

3. $\partial^2 Q_n(\varphi)/\partial\varphi\partial\varphi' - J(\varphi) \xrightarrow{p} 0$, uniformly in φ, with $J(\varphi) = O(1)$, continuous in φ, and $J(\varphi_0)$ is negative definite.

These are sufficient for our purposes, in this paper, and covers standard estimation procedures, such as nonlinear least-squares, instrumental variables, generalized method of moments, or pseudo-maximum likelihood. In addition, the following conditional moment restrictions are satisfied

$$E\left[\varepsilon(W; \varphi_0)|X\right] = 0 \quad a.s. \tag{1}$$

where $\varepsilon(W; \varphi)$ is a $(m \times 1)$ vector of "generalised errors", with typical element $\varepsilon_r(W; \varphi)$, $r = 1, ..., m$. It is possible, for example, that (1) is exploited explicitly in the definition of $Q_n(\varphi)$; see Delgado et al (2006). Whether or not this is the case, all (or a subset of) the conditional moment restrictions of (1) are to be tested after estimation of the specified, or null, model, which is obtained by (erroneously) assuming that $\gamma = 0$. Thus, we imagine that the unknown DGP is a generalisation of the specified null model and that observations on X are available for estimation, and testing, of the null specification.

In order to introduce some necessary notation and additional assumptions, which afford the relatively straightforward investigation of the asymptotic behaviour of smooth kernel-based test statistics, let $K_{ij} = K\left(\dfrac{X_i - X_j}{h}\right)$, $K(.) : \mathbb{R}^k \to \mathbb{R}$ be a kernel function, where $h = h(n)$ is a positive bandwidth parameter. Define $t_{ij}(\varphi) = (\varepsilon(W_i; \varphi) \odot \varepsilon(W_j; \varphi))K_{ij}$, $(m \times 1)$, $i, j = 1, ..., n$, where $\odot$ denotes the Hadamard product (i.e., $t_{ij}(\theta)$ has typical element $\varepsilon_r(W_i; \theta)\varepsilon_r(W_j; \theta)K_{ij}$); $\omega_{rs}(W; \varphi) = \varepsilon_r(W; \varphi)\varepsilon_s(W; \varphi)$, $r, s = 1, ..., m$; and, $C(X; \varphi) = E\left[\varepsilon(W; \varphi)\varepsilon(W; \varphi)'|X\right]$, $(m \times m)$, which has typical element $\upsilon_{rs}(X; \varphi) = E\left[\omega_{rs}(W; \varphi)|\, X\right]$. The CCM test procedures rely on the properties of the following second order degenerate U-Statistic:

$$T_n(\varphi) = \frac{1}{n(n-1)h^k} \sum_{i=1}^{n} \sum_{j \neq i} t_{ij}(\varphi).$$

In order to develop the substantive results, standard assumptions that appear in this literature will be adopted, as described variously in the articles referred to in Section 1 and/or in the U-Statistics literature.

Firstly, certain smoothness conditions are employed, by restricting functions of interest to the following class:

Definition L^α, $\alpha > 0$, is the class of functions $l(.) : \mathbb{R}^k \to \mathbb{R}$ satisfying the following: $\exists$ $\delta > 0$ such that for all $x \in \mathbb{R}^k$, $\sup_{\|d\| \leq \delta} |l(x + d) - l(x)| / \|d\| \leq L(x)$ and $l(.)$ and $L(.)$ have finite moments of order α (or are bounded if $\alpha = +\infty$).

Assumption B

1. $K(.)$ is even, bounded, integrates to 1 and $\lim_{\|u\|\to\infty} \|u\|^k |K(u)| = 0$.

2. $h \to 0$ as $n \to \infty$, such that $nh^k \to \infty$.[3]

3. $f(.) \in L^\infty$.

4. $\varepsilon_r(W;\varphi_0)$, $r = 1, ..., m$, satisfies the following:

   (a) $E[\varepsilon_r(W;\varphi_0)|X] = 0$.
   (b) $E[\omega_{rs}(W;\varphi_0)|X] \in L^4$.
   (c) $E\left[|\varepsilon_r(W;\varphi_0)|^4 |X\right] \in L^2$.

It is assumed, for simplicity of exposition, that $K(.)$ is a symmetric density function, although this is not strictly necessary. Assumption B is sufficient for the following:

$$\Omega_0^{-1/2} nh^{k/2} T_n(\varphi_0) \xrightarrow{d} N(0, I_m), \tag{2}$$

where

$$\Omega_0 = 2E\left[\{C(X;\varphi_0) \odot C(X;\varphi_0)\} f(X)\right] \int K^2(u)\, du$$

has typical element $2E\left[\upsilon_{rs}^2(X;\varphi_0) f(X)\right] \int K^2(u)\, du$, $r, s = 1, ..., m$, and is finite and positive definite. This result is the multivariate generalisation of Hall's (1984) Central Limit Theorem for a second order degenerate U-Statistic.

In addition, the following assumptions are sufficient to justify the asymptotic expansions employed to obtain the limit distribution of the CCM test indicator:[4]

Assumption C

For all $r = 1, ..., m$, and each w, $\varepsilon_r(w;\theta)$ is thrice differentiable in φ, and let
$g_r(W;\varphi) = \partial\varepsilon_r(W;\varphi)\partial\varphi$, $(p \times 1)$, with typical element $\{g_{rt}(W;\varphi)\}$, $t = 1, ..., p$,
$G_r(W;\varphi) = \partial^2\varepsilon_r(W;\varphi)/\partial\varphi\partial\varphi'$, $(p \times p)$,
$F_r(W;\varphi) = \partial vec G_r(W;\varphi)/\partial\varphi$, $(p^2 \times p)$,
satisfying:

1. $E[|\varepsilon_r(W;\varphi_0)| |X] \in L^2$.

2. For all $t = 1, ..., p$: (i) $E[|g_{rt}(W;\varphi_0)|^{8/3}] < \infty$; (ii) $E[g_{rt}(W;\varphi_0)|X] \in L^{8/3}$; (iii) $E[|g_{rt}(W;\varphi_0)| |X] \in L^2$; and, (iv) $E[|g_{rt}(W;\varphi_0)|^2 |X] \in L^2$.

3. (i) $E[\|G_r(W;\varphi_0)\|^{8/3}] < \infty$; and, (ii) $\sup_\varphi \|G_r(W;\varphi)\| < M(W)$, for all r, $E[M(W)] < \infty$ with $\lambda(X) = E[M(W)|X] \in L^2$.

4. (i) $F_r(W;\varphi)$ is continuous in φ, for each w; and, (ii) $\sup_\varphi \|F_r(W;\varphi)\| < P(W)$, for all r, $E[P(W)] < \infty$ with $E[P^2(W)|X] < \infty$.

Some remarks on Assumption C are appropriate. First, rather than the primitive assumptions above, one could appeal to rather high-level assumptions via uniform laws of numbers, as described by Newey (1991, Corollary 4.1) and Newey and Powell (2003, Lemma A2). Second, armed with Assumptions A, B, C1 C2(i),(ii) and C3, and under the true DGP (if it were known), it is then straightforward to show that $\tilde{\Omega}^{-1/2} nh^{k/2} T_n(\tilde{\varphi}) \xrightarrow{d}$

$N(0, I_m)$ where $\tilde{\Omega}$ is any consistent estimator for Ω_0. For example, under some additional dominance assumptions, it can be shown that $\dfrac{2}{n(n-1)h^k} \sum_{i=1}^{n} \sum_{j \neq i} t_{ij}(\tilde{\varphi}) t_{ij}(\tilde{\varphi})'$ will be consistent for Ω_0. Thirdly, the additional restrictions embodied in C2(iii),(iv) and C4 (including thrice differentiability of $\varepsilon_r(W; \varphi)$) are slightly stronger than required in previous analyses of local power. This is partly because (1) is rather general but more importantly because (parametric) *local* generalisations of the specified null model are to be investigated directly, rather than (non-parametric) local failure of the conditional moment condition, (1). This distinction is discussed further in Section 2.3.

Suppose, however, that the researcher does not have knowledge of $Q_n(\varphi)$ and, instead, employs a misspecified a model that erroneously sets $\gamma = 0$, yielding the estimator $\hat{\varphi} = (\hat{\theta}', 0')'$, which solves $\max_\varphi Q_n(\varphi)$ subject to $\gamma = 0$; see Section 2.2. What effect will this have on smooth kernel-based statistics, and associated test procedures, constructed from $T_n(\hat{\varphi})$ and designed to test whether (1) holds, but with $\gamma_0 = 0$? This is the general question that is addressed in the Sections 2.2 and 2.3, below, and the following example illustrates a particular situation that we have in mind:

Example 1 *Suppose $W' = (Y, X')$ and the true DGP is characterised by the regression $E[Y|X] = g(X)$, for some (scalar) function $g(X)$, such that $E[U|X] = 0$ a.s. and $E[U^2|X] = \sigma_0^2$ a.s., where $U = Y - g(X)$. (This generalises the case of omitted variables in which $g(X) = X'\beta_0 + Z'\gamma_0$, where $Z = E[Z^*|X]$ for a vector of "omitted variables" Z^*) However, the researcher performs a linear regression of Y on X and then constructs a CCM test of conditional homoskedasticity. To place this within the context of the preceding discussion, let $z(X) = g(X) - X'\beta_0 \equiv Z$, where $\Pr[|g(X) - X'\beta| > 0] = 1$ for all β. On this basis, define the scalar generalised error $\varepsilon(W; \varphi) = (Y - X'\beta - \gamma Z)^2 - \sigma^2$, so that there exists a φ_0 satisfying $E[\varepsilon(W; \varphi_0)|X] = 0$ a.s., with $\varphi' = (\beta', \sigma^2, \gamma)$. Therefore, hypothetically speaking, this moment restriction affords an asymptotically valid test of conditional homoskedasticity based on $T_n(\tilde{\varphi})$, where $\tilde{\varphi}' = (\tilde{\beta}', \tilde{\sigma}^2, \tilde{\gamma})$ derives from a linear regression of Y on (X', Z), and (moreover) $nh^{k/2} T_n(\tilde{\varphi})$ retains its limit null distribution. In our story, however, the specified (null) model imposes the restriction of $\gamma = 0$ and delivers the estimator $\hat{\theta}' = (\hat{\beta}', \hat{\sigma}^2)$, following a (linear) least squares estimation of Y on X alone, from which the test indicator constructed from $T_n(\hat{\varphi})$ is employed in order to assess the adequacy of the conditional homoskedasticity assumption. The question we wish to address is "how sensitive will this latter procedure be to misspecified regression functional form, as characterised by $\gamma \neq 0$?"*

2.2 The Null DGP and Test Procedure

The true DGP is unknown to the researcher. Rather, the specified null model imposes the restrictions of $H_\gamma : \gamma = 0$ in $Q_n(\varphi)$ giving $Q_n^H(\theta) \equiv Q_n(\varphi_H)$, where $\varphi_H = (\theta', 0')'$. This yields the estimator $\hat{\varphi} = (\hat{\theta}', 0')'$, $\hat{\theta} = \arg\max_\theta Q_n^H(\theta)$, satisfying $\text{plim}_{n\to\infty} \hat{\varphi} = \varphi_* = (\theta_*', 0')'$, and $\theta_* = \theta_0$ if the null model is true. Indeed, by (1), if the null model is true then $E_H[\varepsilon_H(W; \theta_0)|X] = 0$ *a.s.*, where $\varepsilon_H(W; \theta) = \varepsilon(W; \varphi_H)$ and expectations denoted $E_H[.]$, respect the null DGP. In the light of this, the "consistent" kernel-based test statistic is designed to test

$$H_0 : E[\varepsilon_H(W; \theta_*)|X] = 0 \; a.s., \text{ for some } \theta_* \in \Theta, \tag{3}$$

against the (implicit) alternative[5] of

$$H_1 : E[\varepsilon_H(W; \theta)|X] \neq 0 \; a.s., \text{ for all } \theta \in \Theta, \tag{4}$$

rather than the alternative of $\gamma \neq 0$. In order to construct a consistent one-sided procedure, the test is based on the scalar indicator $V_n(\hat{\varphi})$, rather than $T_n(\hat{\varphi})$ directly, where

$$\begin{aligned} V_n(\hat{\varphi}) &= \frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i} v_{ij}(\hat{\varphi}) = T_n(\hat{\varphi})'\iota \\ v_{ij}(\hat{\varphi}) &= \hat{\varepsilon}_i'\hat{\varepsilon}_j K_{ij},\ \hat{\varepsilon}_i = \varepsilon(W_i;\hat{\varphi}) = \varepsilon_H(W_i;\hat{\theta}). \end{aligned} \tag{5}$$

and $\iota = (1, ..., 1)'$ is the $(m \times 1)$ sum vector; see, for example, Delgado et al (2006).

By (1), but under $H_\gamma : \gamma = 0$, (3) holds with $\theta_* = \theta_0$ and Assumptions A-C(i),(ii),C3 imply that $\text{plim}_{n\to\infty} V_n(\hat{\varphi}) = 0$, because (3) is true. Furthermore,

$$nh^{k/2}V_n(\hat{\varphi})/\sqrt{\Sigma_0} \xrightarrow{d} N(0,1) \tag{6}$$

where

$$\Sigma_0 = 2E\left[f(X)\left\|C_H(X;\theta_0)\right\|^2\right]\int K^2(\zeta)\,d\zeta \tag{7}$$

in which $C_H(X;\theta_0) = E_H\left[\varepsilon_H(W;\theta_0)\,\varepsilon_H(W;\theta_0)'\,|X\right]$ and $\|A\|^2 = tr(A'A) = \sum_r\sum_s a_{rs}^2$ for any matrix $A = \{a_{rs}\}$. If $\hat{\Sigma}_n$ denotes any consistent estimator of Σ_0, then the statistic $nh^{k/2}V_n(\hat{\varphi})/\sqrt{\hat{\Sigma}_n} \xrightarrow{d} N(0,1)$, under the null model of $H_\gamma : \gamma = 0$. As noted previously, under some additional dominance conditions, the choice

$$\hat{\Sigma}_n = \frac{2}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i} v_{ij}^2(\hat{\varphi})$$

is consistent for Σ_0 and leads to the asymptotically valid and relatively simple-to-compute statistic

$$\mathcal{J}_n = \frac{\sum_{i=1}^{n}\sum_{j\neq i} v_{ij}(\hat{\varphi})}{\sqrt{2\sum_{i=1}^{n}\sum_{j\neq i} v_{ij}^2(\hat{\varphi})}}$$

A consistent, one-sided test procedure, with nominal significance level $\alpha \times 100\%$, is to reject H_0, in favour of H_1, when $\mathcal{J}_n > c_\alpha$, where c_α satisfies $1 - \Phi(c_\alpha) = \alpha$ and $\Phi(.)$ is the standard normal distribution function. Several studies, however, have found that such critical values are unreliable, because of slow convergence depending upon the choice of bandwidth, h, and number of conditioning variables, X. To try and improve finite sample behaviour, bootstrap procedures could (or should) be employed (tailored to the particular test and null hypothesis under consideration) and these will be described in Section 4.2. [6]

The consistency of the test procedure based on $\mathcal{J}_n$ derives from the fact that, under the true DGP, $\text{plim}_{n\to\infty} V_n(\hat{\varphi}) = E[f(X)\|b(X;\varphi_*)\|^2]$, where $b(X;\varphi_*) = E[\varepsilon(W;\varphi_*)\,|X] = E[\varepsilon_H(W;\theta_*)\,|X]$; thus, whenever (4) holds, $\text{plim}_{n\to\infty} V_n(\hat{\varphi}) = E[f(X)\|b(X;\varphi_*)\|^2] > 0$ implying a that a one-sided test procedure rejects the null for large values of $\mathcal{J}_n$. It is in this sense that the test procedure is a consistent test of (3) against (4). However, this does not ensure that the procedure will be powerful against all sources of null model misspecification. The question is, then, "when might tests based on (5), with (4) as the implicit alternative, be relatively insensitive to other (unconsidered) departures from the specified null model?". Following, for example, Godfrey and Orme (1996) the framework of local alternatives, laid out below, allows an investigation of this.

2.3 Asymptotic Local Analysis

In order to derive the $O_p(1)$ limit distribution of $nh^{k/2}V_n(\hat{\varphi})$ under the true DGP, local generalisations of the null model are considered. Due to the convergence rate of $nh^{k/2}V_n(\hat{\varphi})$, the true (but unknown) DGP is characterised by the Pitman sequence of $\gamma_0 = \delta/\sqrt{nh^{k/2}}$, $0 \leq \|\delta\| < \infty$, $\varphi_0' = (\theta_0', \gamma_0')$, with the specified null model emerging when $\delta = 0$ and where the dependence of φ_0 on n is suppressed for notational convenience. By partitioning δ into separate sub-vectors the impact of a number of unconsidered local alternatives can be identified. However, some care is needed since, for $\delta \neq 0$, the sequence of alternatives converges to the null model at a slower rate than $\sqrt{n}$, implying $\sqrt{n}(\hat{\theta} - \theta_0) \neq O_p(1)$, in general (unlike the situation in Godfrey and Orme, 1996). The following is proved in the Appendix.

Theorem 1 *Define the following:* $J(\varphi_0) = \begin{bmatrix} J_{\theta\theta} & J_{\theta\gamma} \\ J_{\gamma\theta} & J_{\gamma\gamma} \end{bmatrix}$, $\xi = \begin{bmatrix} J_{\theta\theta}^{-1} J_{\theta\gamma}\delta \\ -\delta \end{bmatrix}$, $d(X;\varphi) = E\left[\left. \frac{\partial \varepsilon(W;\varphi)}{\partial \varphi'} \right| X \right]$ *and* $\mu(\varphi_0) = E\left[\|d(X;\varphi_0)\xi\|^2 f(X) \right]$.
Under the true DGP characterised by $\gamma_0 = \delta/\sqrt{nh^{k/2}}$, $0 \leq \|\delta\| < \infty$, *and Assumptions A-C*

$$nh^{k/2}V_n(\hat{\varphi}) \xrightarrow{d} N(\mu_0, \Sigma_0)$$
$$\mathcal{J}_n \xrightarrow{d} N(\Sigma_0^{-1/2}\mu_0, 1)$$

where $\mu_0 = \lim_{n\to\infty} \mu(\varphi_0)$ *and* $\Sigma_0 = \lim_{n\to\infty} \iota'\Omega_0\iota$ *is defined at (7).*

This result articulates the sensitivity of smooth kernel-based test statistics, $nh^{k/2}V_n(\hat{\varphi})$, to local model misspecification. The statistic will be insensitive if, and only if, $\mu_0 = \lim_{n\to\infty} \mu(\varphi_0) = 0$. From Assumption B3, $nh^{k/2}V_n(\hat{\varphi})$ will thus be insensitive to local misspecification if and only if $d(X;\varphi_0)\xi = o(1)$, *a.s.*, for all $\delta \neq 0$.

Remark 1 *Writing* $d_\theta(X;\varphi) = \left[\left. \frac{\partial \varepsilon(W;\varphi)}{\partial \theta'} \right| X \right]$ *and* $d_\gamma(X;\varphi) \left[\left. \frac{\partial \varepsilon(W;\varphi)}{\partial \gamma'} \right| X \right]$, *the sensitivity is decomposed into two parts since* $d(X;\varphi_0)\xi = d_\theta(X;\varphi_0)J_{\theta\theta}^{-1}J_{\theta\gamma}\delta - d_\gamma(X;\varphi_0)\delta$. *This decomposition is reminiscent of that obtained by Godfrey and Orme (1996, equation (17)), with* $d_\gamma(X;\varphi_0)\delta$ *providing the "direct" effect and* $d_\theta(X;\varphi_0)J_{\theta\theta}^{-1}J_{\theta\gamma}\delta$ *the "indirect", or estimation, effect generated by the local inconsistency of* $\hat{\theta}$. *However, the crucial difference is that* $d(X;\varphi_0)$ *is a conditional expectation, not "unconditional" as in Godfrey and Orme's analysis. Because* $\mu_0 = \lim_{n\to\infty} E\left[\|d(X;\varphi_0)\xi\|^2 f(X) \right]$, *it is, of course, exactly this difference which helps deliver the consistency of CCM tests, against certain sources of misspecification, where the corresponding parametric test is inconsistent. Notice also that the distribution of* X *will also effect the sensitivity of the test to non-zero* δ.
Proposition 1, in the Appendix, shows that (in general) $\sqrt{nh^{k/2}}(\hat{\theta} - \theta_0) = O_p(1)$. *However, if* $J_{\theta\gamma} = o(1)$ *then* $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$ *and* $\mu(\varphi_0) = E\left[\|d_\gamma(X;\varphi_0)\delta\|^2 f(X) \right]$, *for all* $\delta \neq 0$.

The proof of Theorem 1, in the Appendix, provides a relatively direct verification of the result. An alternative strategy, under local generalisations of the specified null model, would be: (i) define $\theta_* = \theta_0 + J_{\theta\theta}^{-1}J_{\theta\gamma}\gamma_0$; (ii) expand $nh^{k/2}V_n(\hat{\varphi})$ about θ_* yielding $nh^{k/2}V_n(\hat{\varphi}) = nh^{k/2}V_n(\varphi_*) + o_p(1)$; (iii) show that $nh^{k/2}V_n(\varphi_*) = nh^{k/2}V_n(\varphi_0) + \mu(\varphi_0) + o_p(1)$. Existing analyses of asymptotic local power, for example, in Zheng (1996, Section 4) or Hsiao and Li (2001, Section 3), do not entertain this last stage and thus differ from the approach taken here. These, and other, studies have only been concerned with failure of the CM restriction under test and have therefore taken the null to be $E\left[\varepsilon_H(W;\theta_*)|X\right]=0$, *a.s.*, where $\hat{\theta}-\theta_* \xrightarrow{p} 0$, with local alternatives being $E\left[\varepsilon_H(W;\theta_*)|X\right]=c_n l(X)$, $c_n = O(n^{-1/2}h^{k/4})$ in which $l(X)$ is some unknown function. This obviates the need to address the local inconsistency of $\hat{\theta}$ since, under this sequence, $nh^{k/2}V(\hat{\varphi}) = nh^{k/2}V_n(\varphi_*)+o_p(1)$, where $\varphi_*' = (\theta_*, 0')$. Writing $u=\varepsilon_H(W;\theta_*)-c_n l(X)$, and substituting for $\varepsilon_H(W;\theta_*)$, yields $nh^{k/2}V_n(\varphi_*) \xrightarrow{d} N(\mu_*, \Sigma_0)$, where $\mu_* = E\left[f(X)\left\|l(X)\right\|^2\right]$. Notice that μ_* is unknown, so that this result is silent about the effects of differing sources of local model misspecification which might induce violations of $E\left[\varepsilon_H(W;\theta_*)|X\right]=0$, *a.s.*

The motivation of this paper is quite different in that sources of local model misspecification, which may (or may not) lead to local CM restriction violations, are of concern. In this sense, Theorem 1 refines the asymptotic local analysis summarised in the previous paragraph. It might, therefore, be useful to relate Theorem 1 to existing results in the literature. To do so, consider an expansion of $b(X;\varphi_*)$, as defined above, about $\varphi_*=\varphi_0$. Because $b(X;\varphi_0)=0$, by (1), and imposing sufficient regularity on the conditional distribution of W given X so that $\partial b(X;\varphi_0)/\partial\varphi = E[d(X;\varphi_0)|X]$, we obtain

$$b(X;\varphi_*) = d(X;\varphi_0)\left(\varphi_*-\varphi_0\right)+o(\|\varphi_*-\varphi_0\|).$$

For local model misspecification, and following Kiefer and Skoog (1985), $(\varphi_*-\varphi)$ can be approximated by $\xi/\sqrt{nh^{k/2}}$ and making this substitution gives $b(X;\varphi_*)\cong d(X;\varphi_0)\xi/\sqrt{nh^{k/2}} \equiv l(X)/\sqrt{nh^{k/2}}$, for large n. Thus, *to the same order of approximation,* $\mu_0=\mu_*$, and the result in Theorem 1, reflects the induced local misspecification of the moment condition under test. The following example illustrates:

Example 2 *Consider Zheng's (1996) test for misspecified (regression) functional form, assuming that the null specification is a linear regression of* Y *on* X*. As in Example 1 and in order to exploit Theorem 1, the "true" DGP is expressed as* $E[Y|X]=X'\beta_0+\gamma_0 z(X)$*, where* $z(x)=g(x)-x'\beta_0$ *for some unknown scalar function* $g(.)$*. When* $\gamma_0=0$*, the true DGP is a member of the specified family of models, but when* $\gamma_0\neq 0$ *it is not with, in particular,* $\gamma_0=1$ *yielding* $E[Y|X]=g(X)$*. For* $\gamma_0=\delta/\sqrt{nh^{k/2}}$*, the true DGP describes local generalisations of the specified null model that we wish to exploit. Define* $Q_n(\varphi) = n^{-1}\sum_{i=1}^n (Y_i - X_i'\beta-\gamma Z_i)^2$*,* $Z_i=z(X_i)$*, so that* $\tilde{\theta}-\theta_0 \xrightarrow{p} 0$ *and* $\tilde{\gamma}-\gamma_0 \xrightarrow{p} 0$*, under this "true" DGP. The specified (null) model is obtained by setting* $\gamma=0$*, in* $Q_n(\varphi)$*, yielding the least squares estimator* $\hat{\theta}$*, and* $\theta_*=\theta_0+\Sigma_{xx}^{-1}\Sigma_{xz}\gamma_0 = (1-\gamma_0)\theta_0+\Sigma_{xx}^{-1}\Sigma_{xg}\gamma_0$*, where* $\Sigma_{xx} = E\left[XX'\right]$*,* $\Sigma_{xz}=E\left[z(X)X\right]$ *and* $\Sigma_{xg}=E[g(X)X]$*. Furthermore, if* $\varepsilon(W;\varphi)=Y-X'\theta-\gamma Z$ *and Zheng's test is carried out with* $\varepsilon(W;\hat{\varphi})=\varepsilon_H(W;\hat{\theta})=Y-X'\hat{\theta}$*, then it is easily shown that* $E\left[\varepsilon_H(W;\theta_*)|X\right] = \left(z(X)-X'\Sigma_{xx}^{-1}\Sigma_{xz}\right)\gamma_0 = \left(g(X)-X'\Sigma_{xx}^{-1}\Sigma_{xg}\right)\gamma_0$*. Therefore, under local generalisations of the specified, with* $\gamma_0=\delta/\sqrt{nh^{k/2}}$*,* $0\leq|\delta|<\infty$*, the moment condition under test will also be locally misspecified (to exactly the same order of approximation) as* $E\left[\varepsilon_H(W;\theta_*)|X\right] = n^{-1/2}h^{-k/4}l(X)$ *with* $l(X) = \left(g(X)-X'\Sigma_{xx}^{-1}\Sigma_{xg}\right)\delta = \left(z(X)-X'\Sigma_{xx}^{-1}\Sigma_{xz}\right)\delta$*. This is consistent with Zheng (1996, Section 4) and is also in accord with* $n^{-1/2}h^{-k/4}d(X;\varphi_0)\xi$*, because* $d(X;\varphi_0)=-(X',Z')$ *and* $\xi = \begin{pmatrix}\Sigma_{xx}^{-1}\Sigma_{xz}\delta \\ -\delta\end{pmatrix}$.

3 Applications

As noted in the Introduction, the results of Godfrey and Orme (1996, Example 4.3) show that a parametric test for heteroskedasticity (e.g. Koenker's, 1981, test) is insensitive to local (regression) functional form misspecification (or omitted variables). This is also true of a CCM test for conditional heteroskedasticity, as follows. The scenario is as described in Examples 1 and 2, and define $Q_n(\varphi) = -n^{-1}\sum_{i=1}^n\{\ln\sigma^2 + (Y_i - X_i'\beta - \gamma Z_i)^2/\sigma^2\}$, with $\varphi' = (\beta', \sigma^2, \gamma)$. The specified null model, which imposes $\gamma = 0$, yields $\hat{\theta}' = (\hat{\beta}', \hat{\sigma}^2)$ and $\hat{\varepsilon} = \varepsilon_H(W;\hat{\theta}) = (Y - X'\hat{\beta})^2 - \hat{\sigma}^2 = \hat{U}^2 - \sigma^2$, say, where here (and in what follows) $\hat{U} = Y - X'\hat{\beta}$. An asymptotically valid (scalar) test statistic is

$$\begin{aligned} \mathcal{H}_n &\equiv \frac{\sum_{i=1}^n \sum_{j\neq i} v_{ij}(\hat{\varphi})}{\sqrt{2\sum_{i=1}^n \sum_{j\neq i} v_{ij}^2(\hat{\varphi})}} \\ v_{ij}(\hat{\varphi}) &= (\hat{U}_i^2 - \sigma^2)(\hat{U}_j^2 - \hat{\sigma}^2)K_{ij} \end{aligned} \tag{8}$$

as discussed in Section 2.2. Under these assumptions, and where $\gamma_0 = \delta/\sqrt{nh^{k/2}}$, $0 \leq |\delta| < \infty$, it is straightforward to show that (in the notation of Theorem 1)

$$\begin{aligned} d(X;\varphi_0) &= (0', -1, 0') \\ \xi &= \begin{bmatrix} \Sigma_{xx}^{-1}\Sigma_{xz}\delta \\ 0 \\ -\delta \end{bmatrix} \end{aligned}$$

where Σ_{xx} and Σ_{xz} are defined as before and $d(X;\varphi_0)$ and ξ are partitioned conformably giving $d(X;\varphi_0)\xi = 0$. Thus, an application of Theorem 1 yields $\mu_0 = 0$, so that the CCM test for conditional heteroskedasticity is insensitive to local regression function misspecification, of the order $n^{-1/2}h^{-k/4}$.

An intuitive demonstration of this result runs as follows. With $\gamma_0 \neq 0$, $\hat{\beta} - \beta_* \xrightarrow{p} 0$, where $\beta_* = \beta_0 + \Sigma_{xx}^{-1}\Sigma_{xz}\gamma_0$, and $\hat{\sigma}^2 - \sigma_*^2 \xrightarrow{p} 0$ where $\sigma_*^2 = E[U_*^2]$, $U_* = Y - X'\beta_*$. We can write $U_* = U + (Z - X'\Sigma_{xx}^{-1}\Sigma_{xz})\gamma_0 = U + \tilde{Z}\gamma_0$, say, so that $\sigma_*^2 = \sigma_0^2 + \gamma_0^2 E[\tilde{Z}^2]$, and $E[U_*^2|X] = \sigma_0^2 + \gamma_0^2 E[\tilde{Z}^2|X]$. This yields $E\left[U_*^2 - \sigma_*^2|X\right] = \gamma_0^2(E[\tilde{Z}^2|X] - E[\tilde{Z}^2]) = O(\gamma_0^2)$ which is $o(n^{-1/2}h^{-k/4})$, under the specified local alternatives, and thus not detectable at this order of approximation.

The above conclusion is useful (and extends to the case where, under the null, the conditional variance is parametrically specified as $E[U^2|X] = \tau(X;\eta_0) > 0$) since it suggests a CCM test for misspecified functional form, defined as

$$\begin{aligned} \mathcal{F}_n &\equiv \frac{\sum_{i=1}^n \sum_{j\neq i} v_{ij}(\hat{\varphi})}{\sqrt{2\sum_{i=1}^n \sum_{j\neq i} v_{ij}^2(\hat{\varphi})}} \\ v_{ij}(\hat{\varphi}) &= \hat{U}_i\hat{U}_j K_{ij} \end{aligned} \tag{9}$$

and a CCM test for misspecified conditional heteroskedasticity, $\mathcal{H}_n$ (defined previously), might be carried out separately in order to identify these two possible sources of misspecification (Theorem 1 of Zheng, 1996, ensures that $\mathcal{F}_n$ is robust to conditional heteroskedasticity - local or otherwise). Moreover, asymptotically justified inferences can be obtained from the inspection of the separate tests, $\mathcal{F}_n$ and $\mathcal{H}_n$, respectively, if they are also asymptotically independent (under the null), since then the overall significance level of the (induced) test is simply $1-(1-\alpha_{\mathcal{F}})(1-\alpha_{\mathcal{H}})$, where, for example, $\alpha_{\mathcal{F}}$ is the marginal significance level of $\mathcal{F}_n$. From (2), or the results of Delgado et al (2006), and under the null of $\gamma_0 = 0$, asymptotic independence arises if and only if $E_H\left[U\left(U^2 - \sigma_0^2\right)|X\right] = 0$ or, equivalently $E_H\left[U^3|X\right] = 0$; i.e., conditional symmetry. Thus a test of conditional symmetry would also be a test of asymptotic independence. Another reason for, perhaps, wishing to test the assumption of conditional symmetry of the errors is when a wild bootstrap procedure is to be employed in conjunction with $\mathcal{F}_n$; see Li and Wang (1998). In this case, the (symmetric) wild bootstrap scheme advocated by Davidson and Flachaire (2001,2008), for example, might be expected to provide improved inferences (over other schemes) when the (true) error distribution is also conditional symmetric; see MacKinnon (2006, p.S7).

Following estimation of the specified null model in Example 1, a CCM test of conditional symmetry is based on

$$\begin{aligned}\mathcal{S}_n &\equiv \frac{\sum_{i=1}^n \sum_{j\neq i} v_{ij}(\hat{\varphi})}{\sqrt{2\sum_{i=1}^n \sum_{j\neq i} v_{ij}^2(\hat{\varphi})}} \\ v_{ij}(\hat{\varphi}) &= \hat{U}_i^3 \hat{U}_j^3 K_{ij}.\end{aligned}$$

However, in this case, an application of Theorem 1, with $\varepsilon(W;\varphi) = (Y - X'\beta - \gamma Z)^3$, reveals that $d(X;\varphi_0)\,\xi = 3\sigma_0^2\left(Z - X'\Sigma_{xx}^{-1}\Sigma_{xz}\right)\delta$ under local alternatives $\gamma_0 = \delta/\sqrt{nh^{k/2}}$, $0 \leq |\delta| < \infty$. Thus, whilst (clearly) being insensitive to conditional heteroskedasticity, the limit distribution of $\mathcal{S}_n$ has a non-centrality parameter, μ_0, that is proportional to that of $\mathcal{F}_n$. Thus asymptotic local theory predicts that $\mathcal{S}_n$ will be highly sensitive to local regression function misspecification, unlike the corresponding parametric test for symmetry based on the indicator $\frac{1}{n}\sum_{i=1}^n \hat{U}_i^3$; see Godfrey and Orme (1994).

This feature of the test, it might be argued, is unfortunate in that it could reject conditional symmetry because of regression function misspecification and not asymmetry in the error distribution, *per se*. However, the results in Section 2.3 also provide a possible solution. Rather than $\varepsilon_H(W;\hat{\theta}) = \hat{U}^3$ as the basis for the test, employ $\varepsilon_H(W;\hat{\theta}) = \hat{U}^3 - 3\hat{\sigma}^2\hat{U}$. The so-modified statistic is

$$\begin{aligned}\mathcal{S}_n^M &\equiv \frac{\sum_{i=1}^n \sum_{j\neq i} v_{ij}(\hat{\varphi})}{\sqrt{2\sum_{i=1}^n \sum_{j\neq i} v_{ij}^2(\hat{\varphi})}} \qquad (10) \\ v_{ij}(\hat{\varphi}) &= (\hat{U}_i^3 - 3\hat{\sigma}^2\hat{U}_i)(\hat{U}_j^3 - 3\hat{\sigma}^2\hat{U}_j)K_{ij}.\end{aligned}$$

Clearly, under correct regression functional form, this statistic remains robust to heteroskedasticity. Theorem 1 implies, also, that it will be insensitive to both (locally) misspecified regression functional form and conditional heteroskedasticity, as follows.

The true DGP is now characterised by a simple linear regression model of $Y = X'\beta_0 + \gamma_{10}Z + U$, with Z as before, $E[U|X] = 0$ *a.s.* and $E[U^2|X] = \sigma_0^2 + \gamma_{20}S$, *a.s.*, $S = S(X)$ some scalar function of X. Estimation would employ

$$Q_n(\varphi) = -n^{-1}\sum_{i=1}^n \left\{\ln(\sigma^2 + \gamma_2 S_i) + \left(Y_i - X_i'\beta - \gamma_1 Z_i\right)^2/(\sigma^2 + \gamma_2 S_i)\right\},$$

$S_i = S(X_i)$, with $\varphi' = \left(\beta', \sigma^2, \gamma'\right)$, $\gamma' = (\gamma_1, \gamma_2)$. Here, then,

$$\varepsilon(W;\varphi) = \left(\left(Y - X'\beta - \gamma_1 Z\right)^3 - 3(\sigma^2 + \gamma_2 S)(Y - X'\beta - \gamma_1 Z)\right),$$

so that $E[\varepsilon(W;\varphi_0)|X] = 0$ *a.s.*. The specified null model (imposing $\gamma = 0$) yields $\hat{\varepsilon} = \varepsilon_H(W;\hat{\theta}) = \hat{U}^3 - 3\hat{\sigma}^2\hat{U}$, which is then employed to construct $\mathcal{S}_n^M$ in (10). It immediately follows that

$$d(X;\varphi_0) = (0', 0, 0', 0).$$

Therefore the modified test, (10), is asymptotically insensitive to local misspecification of the regression and error variance functions.

As before, the result may also be substantiated as follows. Again we can write $U_* = U + \tilde{Z}\gamma_{10}$, where, in this case, $\sigma_*^2 = E[U_*^2] = \sigma_0^2 + \gamma_{20}E[S] + \gamma_{10}^2 E[\tilde{Z}^2]$. Hence we have

$$\begin{aligned} E\left[U_*^3 - 3\sigma_*^2 U_*|X\right] &= 3(\sigma_0^2 + \gamma_{20}S)\gamma_{10}E[\tilde{Z}|X] + \gamma_{10}^3 E[(\tilde{Z}^3|X] \\ &\quad -3(\sigma_0^2 + \gamma_{20}E[S] + \gamma_{10}^2 E[\tilde{Z}^2])\gamma_{10}E[\tilde{Z}|X] \\ &= 3\gamma_{10}\gamma_{20}(S - E[S])E[Z|X] + O(|\gamma_{10}|^3). \end{aligned}$$

When the individual misspecifications are $O\left(n^{-1/2}h^{-k/4}\right)$, i.e., $\gamma_0 = O\left(n^{-1/2}h^{-k/4}\right)$, the above conditional expectation is $o(n^{-1/2}h^{-k/4})$ and thus negligible relative to the order of approximation being employed.

In summary, and to the order of approximation considered in this paper, the asymptotic analysis implies the following: a CCM test of correct regression functional form should be robust to heteroskedasticity, but its power (against incorrect regression functional form) may be sensitive to heteroskedasticity and the regressor distribution, although not to the error distribution. The CCM test of conditional homoskedasticity, should be relatively ineffective at detecting incorrect regression functional form and its power against heteroskedasticity may be sensitive to the regressor and error distribution (through the influence of moments of order 4). The modified CCM test of conditional symmetry is robust to heteroskedasticity, under correct regression functional form, but only locally insensitive to both regression misspecification and heteroskedasticity. Its power against asymmetry in the error distribution may be sensitive to the regressor distribution and moments of order 4 and 6 in the error distribution.

4 Monte Carlo Design

4.1 Data Generation

In order to provide some evidence on the quality of the predictions derived from the preceding asymptotic local theory, the actual finite sample behaviour of separate CCM tests (designed to detect incorrect functional form, heteroskedasticity and skewness) are investigated using a Monte Carlo experiment. The design employs generalisations of the following linear regression model

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_3 + \sigma\varepsilon_i = X_i'\beta + \sigma\varepsilon_i, \quad \varepsilon_i \; iid\,(0,1),\; i = 1, ..., n \qquad (11)$$

in which $X_{i1} = 1$, $X_{i3} = \rho X_{i2} + \sqrt{1-\rho^2}V_i$ with X_{i2} and V_i being *either iid* $N(0,1)$ random variables or *iid* standardised χ_2^2 random variables which are combined with ε_i being *either iid* $N(0,1)$ random variables or *iid* standardised χ_2^2 random variables. The specification in (11) is the null model, which is estimated by OLS order that the various tests, defined in Section 4.2 below, can be carried out. There particular (true) parameter values chosen were $\beta_{0j} = 1$, $j = 1,2,3$, $\rho_0 = 0.5$ and $\sigma_0 = 1$, which gives a population R^2 of 75% for (11), and sample sizes considered are $n = 50, 100, 200$. This shall be referred to as DGP_0. However, all the CCM test statistics considered are invariant to $\left(\beta_0', \sigma_0\right)$ when the true DGP is a member of the family described by (11).

Using the same distributional assumptions for $(X_{i1}, X_{i2}, \varepsilon_i)$ and the same values for $(\beta_0', \sigma_0, \rho_0)$, alternative DGPs are defined in the following way:

1. $DGP_1 : Y_i = X_i'\beta_0 + \frac{1}{2}X_{i1}X_{i2} + \sigma_0\varepsilon_i.$
2. $DGP_2 : Y_i = |X_i'\beta_0|^{1/3} + \sigma_0\varepsilon_i.$
3. $DGP_3 : Y_i = (X_i'\beta_0)^{1/3} + \sigma_0\varepsilon_i.$ [7]
4. $DGP_4 : Y_i = X_i'\beta_0 + \sigma_0\sqrt{1 + X_{i1}^2}\varepsilon_i.$
5. $DGP_5 : Y_i = X_i'\beta_0 + \frac{1}{2}X_{i1}X_{i2} + \sigma_0\sqrt{1 + X_{i1}^2}\varepsilon_i.$
6. $DGP_6 : Y_i = |X_i'\beta_0|^{1/3} + \sigma_0\sqrt{1 + X_{i1}^2}\varepsilon_i.$
7. $DGP_7 : Y_i = (X_i'\beta_0)^{1/3} + \sigma_0\sqrt{1 + X_{i1}^2}\varepsilon_i.$

In order to accommodate Theorem 1, within a unified framework, each of the above can be regarded as particular members of the following family, which defines generalisations of (11),

$$Y = X'\beta + \gamma_1 z(X) + \sigma\sqrt{1 + \gamma_2 X_1^2}\varepsilon,$$

where $z(X) = E[Y|X] - X'\beta_0$ and $E[Y|X]$ is specified by $DGP_1 - DGP_7$, respectively; e.g., for DGP_5, $z(X) = X'\beta_0 + \frac{1}{2}X_1X_2$, $\beta = \beta_0$, and $\gamma_1 = \gamma_2 = 1$. All of these $(DGP_1 - DGP_7)$ represent misspecifications of (11): $DGP_1 - DGP_3$ being misspecified regression functional form; DGP_4, neglected heteroskedasticity; and, $DGP_5 - DGP_7$ joint misspecification. Note that under DGP_1, or DGP_5 (misspecified functional form) and multivariate normality for (X_{i1}, X_{i2}), OLS estimation of (11) still yields consistent estimators for the β_j, but not when the regressor values are realisations of χ^2 random variables.

The finite sample behaviour of test statistics, constructed after estimation of (11), is investigated by employing $10,000$ artificial samples of (Y, X_1, X_2) generated according DGP_0, and $5,000$ artificial samples for each of DGP_1 to DGP_7. The results are reported in Section 5.[8]

4.2 Test Statistics

OLS estimation of (11) yields estimators $\hat{\beta}_j$ and residuals $\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \beta_2 X_{i2}$, from which $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n \hat{U}_i^2$. Based on these quantities, each CCM test statistic has the following common structure

$$\mathcal{J}_n = \frac{\sum_{i=1}^n \sum_{j \neq i} \hat{e}_i K_{ij} \hat{e}_j}{\sqrt{2\sum_{i=1}^n \sum_{j \neq i} \hat{e}_i^2 K_{ij}^2 \hat{e}_j^2}} \tag{12}$$

where $K_{ij} = \exp\left(-\frac{1}{2}\sum_{m=1}^2 \left\{\frac{X_{im} - X_{jm}}{h_m}\right\}^2\right)$ and $h_m = \lambda_0 sd(X_j)/n^{1/6}$, where $sd(X_j)$ is the standard deviation of regressor j. However, since $sd(X_j) = 1$, $j = 1, 2$, a common $h = h_m = \lambda_0/n^{1/6}$ is employed with $\lambda_0 = 0.5$; Li and Wang (1998).[9] Sampling experiments were also conducted for $\lambda_0 = 1$ and $\lambda_0 = 2$ to assess sensitivity to choice of bandwidth and these results are summarised in Section 5.

For each separate test statistic, $\hat{e}_i$ is defined as follows:

1. Functional Form: $\hat{e}_i = \hat{U}_i$, yielding $\mathcal{F}_n$.

2. Heteroskedasticity: $\hat{e}_i = \hat{U}_i^2 - \hat{\sigma}^2$, yielding $\mathcal{H}_n$.

3. Skewness: $\hat{e}_i = \hat{U}_i^3$, yielding $\mathcal{S}_n$.

   Two further test statistics are also considered:

4. Since the null hypothesis of the heteroskedasticity test is one of $H_0 : E\left[U^2 - \sigma^2|X\right] = 0$, $a.s.$, this can be exploited yielding the statistic

$$\mathcal{H}_n^K = \frac{\sum_{i=1}^n \sum_{j\neq i} \hat{e}_i K_{ij} \hat{e}_j}{\hat{\tau}_n \sqrt{2\sum_{i=1}^n \sum_{j\neq i} K_{ij}^2}}, \quad \hat{e}_i = \hat{U}_i^2 - \hat{\sigma}^2, \quad \hat{\tau}_n = n^{-1}\sum_{i=1}^n \hat{U}_i^4 - \hat{\sigma}^4$$

   which is rather like a Koenker-type (Studentised) variant of the heteroskedasticity test statistic. [10]

5. $\mathcal{F}_n$ and $\mathcal{S}_n$ are robust to neglected heteroskedasticity. According to the analysis of Section 3, $\mathcal{H}_n$ and $\mathcal{H}_n^K$ should both be "relatively" ineffective in diagnosing misspecified regression function form, but this not so for $\mathcal{S}_n$. In view of this, the modified version of the skewness statistic, as described in Section 3 and denoted $\mathcal{S}_n^M$, is also examined. This employs $\hat{e}_i = \hat{U}_i^3 - 3\hat{\sigma}^2\hat{U}_i$, in (12), and (like $\mathcal{H}_n$ and $\mathcal{H}_n^K$) should be "relatively" insensitive to both misspecified regression functional form and neglected heteroskedasticity.

As noted at the end of Section 2.2, a test procedure which uses these statistics in conjunction with critical values from a standard normal distribution is, in general, unreliable. In order to guard against this, bootstrap, procedures are employed as follows.

4.3 Bootstrap Tests

1. For $\mathcal{H}_n$ and $\mathcal{H}_n^K$ a simple nonparametric (residual resampling) bootstrap procedure (denoted BS) is employed, which respects the test's null hypothesis of homoskedasticity. The resulting bootstrap test procedures shall also be denoted $\mathcal{H}_n$ and $\mathcal{H}_n^K$, respectively. Specifically, B artificial samples of size n are generated from

$$Y_i^* = \hat{\beta}_0 + X_i'\hat{\beta} + U_i^*, \quad i = 1, \ldots, n, \tag{13}$$

   where $U_1^*, U_2^*, \ldots, U_n^*$ is a random sample drawn with replacement from $\hat{U}_1, \hat{U}_2, ..., \hat{U}_n$, and for each artificial sample the CCM heteroskedasticity test statistics, described above, are constructed. In each case, if these artificially generated test statistics are denoted by $\mathcal{J}_{n1}^*, \mathcal{J}_{n2}^*, \ldots, \mathcal{J}_{nB}^*$, the p-value of $\mathcal{J}_n$ ($= \mathcal{H}_n$, or $\mathcal{H}_n^K$) can be estimated by[11]

$$PV_{BS} = \frac{\sum_{b=1}^B \mathbf{1}\left(\mathcal{J}_{nb}^* > \mathcal{J}_n\right)}{B}.$$

   The null hypothesis (of homoskedasticity) is then rejected when $PV_{BS} \leq \alpha$, where α is the desired significance level.

   Since the test statistics under consideration are asymptotic pivots, this bootstrap test procedure delivers asymptotically valid inferences but has an Error in Rejection Probability (ERP) which is of smaller order in n than that of the asymptotically valid test procedure which employs $\Phi(.)$ as the reference distribution; see Beran (1988).

Although not pursued here, if an auxiliary assumption of normality for the ε_i in (11) is imposed, then the Monte Carlo methods of Dufour, Khalaf, Bernard and Genest (2004) can be employed to obtain "exact" finite sample inferences by generating the U_i^* in (13) from a standard normal distribution. However, if the normality assumption is wrong, this yields an ERP which is of the same order in n as that of the asymptotically valid test procedure which employs $\Phi(.)$ as the reference distribution; see Godfrey, Orme and Santos-Silva (2006) for an analysis of this in the context of parametric tests for heteroskedasticity.

2. In order to maintain some robustness to heteroskedasticity, the test procedures for each of $\mathcal{F}_n$, $\mathcal{S}_n$ and $\mathcal{S}_n^M$ employ a wild bootstrap scheme (denoted WBS) using the method of Davidson and Flachaire (2008); see Li and Wang (1998) for a detailed justification of the wild bootstrap in this context. The resulting test procedures shall also be denoted $\mathcal{F}_n$, $\mathcal{S}_n$ and $\mathcal{S}_n^M$, respectively. Specifically, the required B artificial samples of size n are generated from

$$Y_i^* = \hat{\beta}_0 + X_i'\hat{\beta} + \hat{U}_i\eta_i, \quad i = 1, \ldots, n, \tag{14}$$

where η_i are *iid* and symmetric with two-point discrete distribution defined by $\Pr(\eta_i = -1) = \Pr(\eta_i = 1) = 0.5$. The artificially generated test statistics are denoted by $\mathcal{J}_{n1}^*, \mathcal{J}_{n2}^*, \ldots, \mathcal{J}_{nB}^*$, and the p-value of $\mathcal{J}_n$ can be estimated by

$$PV_{WBS} = \frac{\sum_{b=1}^{B} \mathbf{1}\left(\mathcal{J}_{nb}^* > \mathcal{J}_n\right)}{B}.$$

The relevant null hypothesis (correct regression functional form or conditional symmetry) is then rejected when $PV_{WBS} \leq \alpha$, where α is the desired significance level.

This symmetric wild bootstrap scheme respects a null hypothesis of error symmetry for $\mathcal{S}_n$ and $\mathcal{S}_n^M$ and is thus an appropriate resampling scheme in this case. Symmetry is not, however, part of the null hypothesis for $\mathcal{F}_n$ but, even so (and although Beran's arguments are not applicable here), this procedure will deliver asymptotically valid inferences (in both cases) with an ERP which is of the same order in n as that of the asymptotically valid test procedure which employs $\Phi(.)$ as the reference distribution. On the other hand, and as noted previously, when the true errors are also conditionally symmetric one might expect improved inferences from employing this, rather than any other, wild bootstrap scheme and, more generally, Davidson and Flachaire (2008, p.169) conclude that "in most practical contexts [whether of not the true error distribution is symmetric], use of [this] wild bootstrap with constrained residuals should provide satisfactory inference" and the evidence provided by Davidson, Monticini and Peel (2007), is also in agreement with this.

For each of the above two bootstrap procedures the number of bootstrap samples generated was $B = 400$ across all DGPs.

5 Monte Carlo Results

5.1 Homoskedastic errors

[Insert Table 1 about here]

Table 1 reports the rejection frequencies (in percentages) for each of the test procedures under DGP_0 at nominal significance levels of 10%, 5% and 1%. These are based on $10,000$ replications of sample data and $B = 400$ bootstrap samples for each of the procedures. The top half of the Table summarises the results for Normal errors, combined with either Normal or Chi-Square regressors. Here, the regression function is correctly specified and the errors are homoskedastic and symmetric, thus all entries are estimated (actual) significance levels. In this case, all procedures deliver estimated significance levels which are in close agreement with the desired nominal levels, for both sets of regressors. Like Li and Wang (1998), since we have tailored the bootstrap to the specific parametric model under test (rather than employing a more general asymptotically valid bootstrap scheme), we get much better results than those reported by Delgado et al (2006). The lower half of Table 1 reports rejection frequencies under (asymmetric) Chi-Square errors, again combined with either Normal or Chi-Square regressors. In this case, the entries for $\mathcal{F}_n$, $\mathcal{H}_n$ and $\mathcal{H}_n^K$ are in accord with asymptotic theory, being estimated significance levels which are very close to the desired nominal levels across both sets of regressors - these procedures should be asymptotically insensitive to asymmetry of the errors. The entries for $\mathcal{S}_n$ and $\mathcal{S}_n^M$ are estimated powers (which increase with n). These procedures exhibit relatively higher power under asymmetric regressors and (interestingly) $\mathcal{S}_n^M$ is the more powerful, of the two, under both regressor distributions; see, also, the discussion of Table 2b below.

Tables 2 reports the rejection frequencies of all test procedures, at the 5% nominal significance level only, when there is misspecified regression functional form, under both Normal and Chi-Square, homoskedastic, errors. These are based on $5,000$ replications of sample data and $B = 400$ bootstrap samples. Firstly, under $DGP_1 - DGP_3$ (misspecified regression functional form, homoskedasticity and possibly asymmetric errors) in Table 2, the entries for $\mathcal{F}_n$ are estimated powers, which increase with n and which are relatively stable across the two error distributions, for a given set of regressors. However, although asymptotically negligible for parametric tests, the direct influence of the regressor density on the $O(1)$ sampling behaviour of this CCM tests is self-evident from Theorem 1 and is reflected in Table 2 in that, for a given error distribution, $\mathcal{F}_n$ is more powerful when the regressors are Chi-Squared, rather than Normal. Furthermore, it is most powerful at detecting DGP_3 and least powerful at detecting DGP_2. The top half of Table 2 (normal errors) also illustrates, vividly, the lack of robustness of the (unmodified) CCM test of conditional symmetry, to misspecified regression functional form: for both sets of regressors: consistent with the prediction of Section 3, the rejections rates of $\mathcal{S}_n$ are only slightly less than the power estimates obtained for $\mathcal{F}_n$.

[Insert Table 2 about here]

5.1.1 Misspecified Regression Functional Form

To illustrate the level of sensitivity/insensitivity that the remaining procedures exhibit to (unconsidered) misspecified regression functional form, Table 2a reports the increases in rejection rates (to the nearest integer), of all the test procedures, under $DGP_1 - DGP_3$

(relative to DGP_0) employing $\mathcal{F}_n$ as benchmark; that is, the difference between the corresponding entries in Tables 1 and 2, with that for $\mathcal{F}_n$ providing a guide to the severity of the perceived misspecification. Firstly, and as noted above, the increase in rejection frequencies reported in Table 2a show that $\mathcal{F}_n$ has relatively more power under Chi-Square regressors. Second, $\mathcal{H}_n$ is relatively ineffective (compared with $\mathcal{F}_n$), under each of $DGP_1 - DGP_3$, with $\mathcal{H}_n^K$ exhibiting remarkable insensitivity under both symmetric and asymmetric regressors. For example, under DGP_3, Normal regressors and errors, the increase in the rejection rate of $\mathcal{F}_n$ is 48, at $n = 100$, whilst that of $\mathcal{H}_n$ is (virtually) 0. However, $\mathcal{H}_n$ is more sensitive to DGP_1 : with Chi-Square regressors and Normal errors the corresponding increases are 61 and 15, respectively, but the latter (for $\mathcal{H}_n$) is still only one-quarter of that for $\mathcal{F}_n$. Across all regressor/error configurations $\mathcal{H}_n$ and $\mathcal{H}_n^K$ are quite insensitive to DGP_2 and DGP_3, with rejection frequencies very close to 5% when compared with the power of the $\mathcal{F}_n$ procedure. Turning, now, to $\mathcal{S}_n^M$ the results in Table 2a indicate that it is fairly insensitive under Normal errors (in accord with the predictions of Section 2.3) with *increases* in rejection frequencies (over those in Table 1) being 18 at most (which occurs with Chi-Square regressors and $n = 200$, compared with an increase of 88 for $\mathcal{F}_n$) but much less than this elsewhere. The increases for $\mathcal{S}_n$, under Normal errors, again reflect the fact it is not (locally) robust to misspecified regression functional form. Finally, and for completeness, the increase in rejection rates of $\mathcal{S}_n^M$ are also provided under Chi-Square errors (in the bottom half of Table 2a), and some of these are relatively large and negative suggesting that, although locally robust under normality, misspecified regression functional form can reduce the power of this procedure to detect skewness.

[Insert Table 2a about here]

5.1.2 Asymmetric Errors

The sensitivity of all procedures to asymmetric errors, and in particular, $\mathcal{S}_n$ and $\mathcal{S}_n^M$, is summarised in Table 2b which provides the increases in the rejections frequencies under Chi-square errors, over those obtained under Normal errors, for each of $DGP_0 - DGP_3$. The entries for $\mathcal{S}_n^M$ in Table 2b can be used as a guide to the perceived severity of the error distribution asymmetry. The entries under DGP_0 illustrate the (previously) noted robustness of $\mathcal{F}_n$, $\mathcal{H}_n$ and $\mathcal{H}_n^K$ to asymmetry of the error distribution, and also the relatively higher power of $\mathcal{S}_n^M$ in detecting such asymmetry. The entries under $DGP_1 - DGP_3$ also provide evidence on the relatively benign effect that asymmetry still has on the sampling behaviour of $\mathcal{F}_n$, $\mathcal{H}_n$ and $\mathcal{H}_n^K$ under misspecified regression functional form. However, under any given $DGP_1 - DGP_3$ and regressor combination, the increase in rejection frequencies for $\mathcal{S}_n$ is always less than that for $\mathcal{S}_n^M$, indicating that the latter is more powerful a detecting asymmetric errors.[12] For example, focussing on $n = 100$, and under Normal regressors, the changes in rejection frequencies of $\mathcal{S}_n$ are (approximately) -13, 5 and 19, across DGP_1, DGP_2 and DGP_3, respectively. The corresponding changes for $\mathcal{S}_n^M$ are 64, 64 and 25. For Chi-Square regressors, these are 22, 28 and -18, for $\mathcal{S}_n$, and 30, 71 and 56 for $\mathcal{S}_n^M$. Thus, under misspecified regression functional form, not only can the rejection rate of $\mathcal{S}_n$ actually fall when the errors change from Normal to Chi-Square, but it is always dominated by $\mathcal{S}_n^M$ which (as a consequence of its asymptotic local robustness to misspecified regression functional form) always exhibits positive power. Under Normal regressors, the power of $\mathcal{S}_n^M$ is unaffected by DGP_1 and DGP_2, but is under DGP_3. Under Chi-Square regressors, its power is not unduly affected by DGP_2, but is by DGP_1 and DGP_3.

[Insert Table 2b about here]

5.2 Heteroskedastic errors

Table 3 reports the rejection frequencies for each of the test procedures under DGP_4 (heteroskedasticity and, possibly, asymmetric errors) and $DGP_5 - DGP_7$ (misspecified regression functional form, heteroskedasticity and possibly asymmetric errors); again, at the 5% nominal significance level, based on $5,000$ replications of sample data and $B = 400$ bootstrap samples. Under DGP_4, the entries for $\mathcal{H}_n$ and $\mathcal{H}_n^K$ are powers and show that $\mathcal{H}_n$ is much more powerful (at detecting heteroskedasticity) than $\mathcal{H}_n^K$ and both are less powerful under Chi-Square errors. This latter result is to be expected since Chi-Square errors will inflate the asymptotic variance of the numerator in $\mathcal{H}_n$. However, (interestingly) the regressor distribution has negligible impact on power. As expected, the entries for $\mathcal{F}_n$ under DGP_4 show that it is quite robust to heteroskedasticity and skewness of the errors, but marginally less so with Chi-Square regressors. Similarly, $\mathcal{S}_N$ and $\mathcal{S}_n^M$ are robust to heteroskedasticity, under Normal errors.

[Insert Table 3 about here]

5.2.1 Misspecified Regression Functional Form

By comparing the entries under $DGP_5 - DGP_7$, with those under DGP_4, the effect of regression specification error can be gauged and, as with Table 2a in the homoskedastic case, this is facilitated by Table 3a which reports the increases in rejection frequencies for the heteroskedastic case. Here, and similar to Table 2a, the entries for $\mathcal{H}_n$ reveal that its behaviour remains fairly insensitive to DGP_6 and DGP_7, and only mildly so to DGP_5; i.e., its power to detect heteroskedasticity is largely undiminished. Note, though, that a comparison of the entries for $\mathcal{F}_n$ in Table 3a with those in Table 2a show that $\mathcal{F}_n$ is less powerful under this form of heteroskedasticity than with homoskedastic errors (the increases reported in Table 3a are less than those in Table 2a). This is to be expected, since (relative to the homoskedastic case) it is easily seen that the variance in the limit distribution of the test indicator will be larger. (A similar result holds for $\mathcal{S}_n^M$: compare the rejection frequencies under DGP_4 and Chi-Square errors in Table 3 with those in Table 1.) As in the homoskedastic case, and as expected, $\mathcal{F}_n$ is more powerful with Chi-Square regressors. This is in direct contrast to the previously noted behaviour of $\mathcal{H}_n$, which remains relatively insensitive to the regressor distribution but which is less powerful under Chi-Square errors.

Turning now to the skewness tests, and whereas the power of $\mathcal{H}_n$ is not overly influenced by $DGP_5 - DGP_7$, the power of $\mathcal{S}_n^M$ to detect asymmetry in the errors (as with homoskedastic errors) is particularly sensitive to DGP_7, with both sets of regressors, and DGP_5 with Chi-Square regressors. Moreover, under the null of symmetry, comparing the entries in the top half of Table 3a with those in Table 2a show that $\mathcal{S}_n^M$ is less robust to misspecified regression functional form under heteroskedastic errors.

[Insert Table 3a about here]

5.2.2 Asymmetric Errors

Table 3b summarises the effect of asymmetric errors by reporting the changes in rejection frequencies under Chi-Square errors, relative to those obtained under Normal errors for each of $DGP_4 - DGP_7$. Under DGP_4 this shows, as in the homoskedastic case, that $\mathcal{S}_n^M$ is more powerful than $\mathcal{S}_n$, and (by comparing these with the entries under DGP in Table 2b) confirms that both are less powerful under heteroskedastic errors. Comparing entries for

$\mathcal{S}_n^M$ under $DGP_5 - DGP_7$, with those under DGP_4, shows the sensitivity of the modified skewness test. The behaviour of $\mathcal{F}_n$ is not overly affected by the error distribution, whilst the negative entries for $\mathcal{H}_n$ confirm that it is less powerful under Chi-Square errors. The sensitivity of the power of $\mathcal{S}_n^M$ has the same qualitative features as those found under homoskedastic errors: it is particularly sensitive to DGP_7, with both sets of regressors, and DGP_5 with Chi-Square regressors.

[Insert Table 3b about here]

5.3 Summary

The main features to emerge from the Monte Carlo experiments, concerning the sensitivity of the tests procedures considered, can be summarised as follows:

1. The sampling behaviour of $\mathcal{F}_n$ is very good under the null of correct regression functional form, being robust to heteroskedasticity and asymmetry of the errors. Although not affected by asymmetry of the error distribution, its power is affected by heteroskedasticity and/or the distribution of the regressors (as $O(1)$ theory predicts).

2. $\mathcal{H}_n^K$ is remarkably robust to regression function misspecification, but lacks power to detect heteroskedasticity relative to $\mathcal{H}_n$. The latter is less robust, but still quite insensitive even when $\mathcal{F}_n$ suggests quite severe (non-local) regression function misspecification. The power of $\mathcal{H}_n$ to detect heteroskedasticity does not appear to be unduly sensitive the regressor distribution, in the experiments considered here, but it is sensitive to the error distribution (as $O(1)$ theory predicts).

3. The skewness test, $\mathcal{S}_n$, is robust to heteroskedasticity but very sensitive to misspecified regression functional form. Under correct regression function specification, it is much less powerful at detecting asymmetry than the modified procedure, $\mathcal{S}_n^M$ (which is also robust to heteroskedasticity) under both homoskedastic and heteroskedastic errors and across both regressor distributions. S_n^M is relatively insensitive to regression function misspecification, (compared with $\mathcal{S}_n$ and $\mathcal{F}_n$) under symmetric errors, but less so under heteroskedasticity.

The above summarise the results when the bandwidth was $h = \lambda_0/n^{1/6}$, with $\lambda_0 = 0.5$. Some experiments were also conducted with $\lambda_0 = 1$ and $\lambda_0 = 2$, to assess the influence of bandwidth selection.[13] Although previous studies (e.g., Zheng, 1996) have indicated that bandwidth selection can adversely affect estimated significance levels when asymptotic critical values are employed, this appears not to be the case with the bootstrap procedures employed here. Li & Wang (1998) reported the same findings but for a wild bootstrap procedure with $\mathcal{F}_n$ only. Generally, all procedures have estimated significance levels close to the desired level for all three values of λ_0; although $\mathcal{F}_n$ can be a little oversized under heteroskedasticity (DGP_4) when $\lambda_0 = 1$ or 2. On the other hand, under misspecification the sensitivity of the various procedures generally increase with λ_0. This is, however, explicable given Theorem 1. As λ_0 increases, so does h for any given n. But given that the DGPs are fixed in the Monte Carlo experiments, the effect is that the perceived misspecification becomes "less local", so sensitivity increases.

6 Conclusion

The theoretical literature provides applied workers with, potentially, an array of Consistent Conditional Moment (CCM) tests (joint and/or separate) which might be employed, post estimation, to assess the validity of particular CM restrictions explicit in, or implied by, an assumed parametric model; e.g., correct regression functional form, conditional homoskedasticity, conditional information matrix equalities, etc. Each CCM procedure is designed to be consistent against any model misspecification which induces failure of the CM restriction(s) under test. This paper contributes to this literature by providing a general theoretical treatment which describes the asymptotic behaviour of separate CMM tests under various sources of local model misspecification; i.e., local generalisations of the specified (or null) model. Firstly this unifies and refines existing results. Secondly, it provides a basis for considering whether CCM test procedures might be constructively employed in a multiple comparison procedure, of the type discussed by Bera and Jarque (1982). Thirdly, the general formulae provided afford an examination of any new procedure that can be interpreted as a CCM test; for example, the test of conditional symmetry considered in this paper. Fourthly, and similarly in spirt to Bera and Yoon (1993), Bera, Montes-Rojas and Sosa-Escudero (2009), the results suggest how a CCM test might be modified in order to reduce its sensitivity to sources of misspecification it is not designed to test.

As in Delgado et al (2006), an particular example explored in this paper is a model where the dependent variable has a parametrically specified conditional mean and conditional variance. Two separate CCM tests, for each part of this model specification (regression and conditional variance functional form), have been developed in the literature. Whilst, by construction, the CCM test of regression functional form is robust to conditional heteroskedasticity, the application of Theorem 1 in this paper shows that the CCM test of conditional variance functional form is asymptotically ineffective in detecting (local) regression functional form misspecification. These two CCM tests are also asymptotically independent under conditional symmetry (of the dependent variable). Moreover, the general results in this paper indicate how a CCM test of conditional symmetry can be constructed, but which remains asymptotically robust to (local) regression and conditional variance misspecification. Such a test might be employed to test the assumption of asymptotic independence between the CCM test of regression functional form and that of conditional variance functional form.

Against fixed (i.e., non-local) regression function specification error, one would expect the various CCM tests considered, here, to consistent. But this offers no useful guidance, to applied workers, on the relative performance of individual tests in a practical (finite sample) situation. As a possible way forward, the asymptotic local analysis employed in this paper provides a theoretical approximation to the sampling behaviour for the various tests when misspecification is not overwhelming. A such, one can interpret the results obtained as follows: (a) relative to the (correct) regression functional form CCM test, one can expect the conditional variance and modified conditional symmetry CCM tests to be quite ineffective at detecting regression specification error, but not so the unmodified conditional symmetry test which might be nearly as powerful as the regression functional form CCM test; (b) the regression functional form CCM test and heteroskedasticity CCM test should be robust to skewness (local or otherwise) under their respective nulls; and, (c) the regression functional form CCM test and modified-skewness CCM test should be robust to heteroskedasticity (local of otherwise) under their respective nulls.

Broadly speaking, the Monte Carlo evidence supports these predictions, but also uncovers other interesting behaviour. In particular, (and as predicted by the $O(1)$ asymptotic theory) the power of the functional form CCM test is insensitive to the error distribution but is sensitive to the regression distribution. On the other hand, asymptotic theory predicts that the power of the heteroskedasticity test should be sensitive to the error distribution (through kurtosis) as well as the regressor distribution, although our Monte Carlo experiments suggest relative insensitivity to the latter. Furthermore, the raw (unstudentised) version of the heteroskedasticity test, $\mathcal{H}_n$, is more powerful than the studentised version $\mathcal{H}_n^k$. A further theoretical examination of the sensitivity/insensitivity of these tests to the distribution of the regressors might prove useful, but is beyond the scope of the current paper.

Finally, the asymptotic analysis employs local alternatives which converge to the null at the rate $n^{1/2}h^{k/4}$. This follows most of the related work in the literature and, in so doing, affords a unification of existing results. However, future research could generalize this method to other type of alternatives of interest (e.g., high frequency local alternatives as in Fan and Li, 2000) and other testing methodologies (such as those based on integral transforms, as in Whang, 2001).

Notes

[1] As, for example, advocated by Bera and Yoon (1993) and Bera, Montes-Rojas and Sosa-Escudero (2009), for parametric tests. Indeed, Zheng's (1996) CCM for regression functional form misspecification is, by construction, robust to heteroskedasticity.

[2] The term *unconsidered local alternative* was coined by Godfrey and Orme (1996) in order to refer to any alternative hypothesis locally distinct from that *implicit alternative* against which a particular Conditional Moment Test has highest power (see Davidson and MacKinnon, 1987).

[3] See Silverman (1986). For ease of exposition attention is restricted to deterministic h. Data driven bandwidths are often used in practice, but their use does not change the conclusions of this paper (at least to a first order approximation); see, for example, Delgado et al (2006).

[4] Although the focus here is on independent data, primitive assumptions can also be stated for the weakly dependent data case which yield, effectively, the same first order results, provided $\varepsilon(W;\theta_0)$ is a martingale difference; see, for example, Fan and Li (1999) and Hsiao and Li (2001).

[5] See Davidson and MacKinnon (1987, p.1305).

[6] Delgado et al (2006) proposed a generally applicable bootstrap scheme which exhibited varaible quality in their sampling experiments. However, for the tests considered in our Monte Carlo experiments more familiar bootstrap schemese can be applied which seem to yield quite close aggreement bewteen desired and actual significance levels.

[7] Where $x^{1/3} \equiv s(x)\,|x|^{1/3}$, with $s(x) = 1$, $x > 0$, and $s(x) = -1$, $x \leq 0$.

[8] Results were obtained using GNU Octave v.3.0.3. Copyright © 2008 John W. Eaton and others.

[9] To be consistent with previous notation, K_{ij} should be devided by h^2, however this is irrelevant when constructing the test statistic.

[10] This does assume that the regression model errors are also homokurtic.

[11] See Broman and Caffo (2003).

[12] This characteristic could also be illustrated by reporting rejection frequencies that employ empirical critical values obtained from the sampling experiments that provided the entries for the top half of Table 2.

[13] Details are available from the authors upon request.

References

[1] Bera, A.K. and C.M. Jarque (1982). Model Specification Tests: A Simultaneous Approach. *Journal of Econometrics,* 20, 59-82.

[2] Bera, A.K. G. Montes-Rojas and W. Sosa-Escudero (2009). Testing Under Local Misspecification and Artificial Regressions. *Economics Letters,* forthcoming.

[3] Bera, A.K. and M.J. Yoon (1993). Specification Testing with Locally Misspecified Alternatives. *Econometric Theory,* 9, 649-658.

[4] Beran, R. (1988). Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements. *Journal of the American Statistical Association,* 83, 687–97.

[5] Bierens, H. (1982). Consistent Model Specification Tests. *Journal of Econometrics*, 20, 105-134.

[6] Bierens, H. (1990). A Consistent Conditional Moment Test of Functional Form. *Econometrica*, 58, 1443-1458.

[7] Bierens, H. and W. Ploeberger (1997). Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica*, 65, 1129-1151.

[8] Broman, K. W. and B. S. Caffo (2003). Simulation-Based P-Values: Response to North et al. *American Journal of Human Genetics*, 72, 496.

[9] Davidson, J., A. Monticini and Peel, D. (2007). Implementing the Wild Bootstrap Using a Two-point Distribution. *Economics Letters*, 96, 309-315.

[10] Davidson, R. and Flachaire, E. (2001). The Wild Bootstrap, Tamed at Last. Queen's Economics DepartmentWorking Paper No. 1000, Queen's University, Kingston,Canada.

[11] Davidson, R. and Flachaire, E. (2008). The Wild Bootstrap, Tamed at Last. *Journal of Econometrics*, 146, 162-169.

[12] Davidson, R. and J.G. MacKinnon (1987). Implicit Alternatives and the Local Power of Test Statistics. *Econometrica,* 55, 1305-1329.

[13] Delgado, M.A., M.A. Dominguez and P. Lavergne (2006). Consistent Tests of Conditional Moment Restrictions. *Annales d'Économie et de Statistique,* 81, 33-67.

[14] Dufour, J.-M., L. Khalaf, J.-T. Bernard and I. Genest (2004). Simulation-Based Finite-Sample Tests for Heteroskedasticity and ARCH Effects. *Journal of Econometrics,* 122, 317–47.

[15] Ellison, G. and S. Fisher Ellison (2000). A Simple Framework for Nonparametric Specification Testing. *Journal of Econometrics,* 96, 1-23.

[16] Eubank, R.L. and C.H. Spiegelman (1990). Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques. *Journal of the American Statistical Association*, 85, 387-392.

[17] Fan, Y. and Q. Li (1996). Consistent Model Specification Tests: Omitted Variables, Parametric and Semi-Parametric Functional Forms. *Econometrica*, 64, 865-890.

[18] Godfrey, L.G. and C.D. Orme (1994). The Sensitivity of Some General Checks to Omitted Variables in the Linear Model. *International Economic Review*, 35, 489-506.

[19] Godfrey, L.G. and C. D. Orme (1996). On the Behaviour of Conditional Moment Tests in the Presence of Unconsidered Local Alternatives. *International Economic Review*, 37, 263-281.

[20] Godfrey, L.G., C.D. Orme, C.D. and J.M.C. Santos-Silva, (2006). Simulation-Based Tests for Heteroskedasticity in Linear Regression Models: Some Further Results. *Econometrics Journal*, 9, 76-97.

[21] Hall, P. (1984). Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. *Journal of Multivariate Analysis*, 14, 1-16.

[22] Härdle, W. and E. Mammen (1993). Comparing Nonparametric versus Parametric Regression Fits. *Annals of Statistics*, 21, 1926-1947.

[23] Hong, &. (1993). Consistent Testing for Heteroskedasticity of Unknown Form. Manuscript, University of California, San Diego.

[24] Hsiao, C. and Q. Li (2001). A Consistent Test for Conditional Heteroskedasticity in Time Series Models. *Econometric Theory*, 17, 188-221.

[25] Kiefer, N.M. and G.R. Skoog (1985). Local Asymptotic Specification Error Analysis. *Econometrica*, 52, 873-885.

[26] Koenker, R. (1981). A Note on Studentizing a Test for Heteroscedasticity. *Journal of Econometrics,* 17, 107-112.

[27] Li, Q. and S. Wang (1998). A Simple Consistent Bootstrap Test for a Parametric Regression Function. *Journal of Econometrics*, 87, 145-165.

[28] MacKinnon, J.G. (2006). Bootstrap Methods in Econometrics. *Economic Record,* 82, S2-S18.

[29] Newey, W.K. (1991). Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometric* 59, 1161-1167.

[30] Newey, W.K. and J.L Powell (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica* 71, 1565–1578.

[31] Powell, J.L., J.H. Stock, and T.M. Stoker (1989). Semiparametric Estimation of Index Coefficients. *Econometrica* 57, 1403 1430.

[32] Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

[33] Stinchcombe, M.B. and H. White (1998). Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative. *Econometric Theory*, 14, 295-325.

[34] Stute, W. (1997). Nonparametric Model Checks for Regression. *Annals of Statistics*, 25, 613-641.

[35] Whang, Y-J. (2000). Consistent Bootstrap Tests of Parametric Regression Function. *Journal of Econometrics*, 98, 27-46.

[36] Whang, Y-J. (2001). Consistent specification testing for conditional moment restrictions. *Economics Letters,* 71, 299-306.

[37] Zheng, J.X. (1994). A Consistent Nonparametric Test of Heteroskedasticity. Manuscript, University of Texas, Austin.

[38] Zheng, J.X. (1996). A Consistent Test of Functional Form via Nonparametric Estimation Techniques. *Journal of Econometrics*, 75, 263-289.

A Proof of Theorem 1

First, a description of the asymptotic behaviour of $(\hat{\theta} - \theta_0)$ will be required. As noted, and unlike the situation in standard parametric inference where the Pitman sequence is $O(n^{-1/2})$, in this case $\sqrt{n}(\hat{\theta} - \theta_0)$ is not $O_p(1)$, in general. Unless stated otherwise, definitions are as given in the main text:

Proposition 1 *Under Assumption A, B2 and the local sequence of alternatives, $\gamma_0 = \delta/\sqrt{nh^{k/2}}$, $0 \leq \|\delta\| < \infty$,*

$$\sqrt{nh^{k/2}}\,(\hat{\varphi} - \varphi_0) = \begin{pmatrix} J_{\theta\theta}^{-1} J_{\theta\gamma}\delta \\ -\delta \end{pmatrix} + o_p(1).$$

Proof. *A mean value expansion of $\sqrt{nh^{k/2}}\partial Q_n(\hat{\varphi})/\partial\theta = 0$ about $\hat{\varphi} = \varphi_0$ yields*

$$0 = \sqrt{nh^{k/2}}\partial Q_n(\varphi_0)/\partial\theta + \left[\partial^2 Q_n(\bar{\varphi})/\partial\theta\partial\theta'\right]\sqrt{nh^{k/2}}(\hat{\theta} - \theta_0) - \left[\partial^2 Q_n(\bar{\varphi})/\partial\theta\partial\gamma'\right]\delta,$$

where $\bar{\varphi}$ is the usual mean value (which may be different for each row of $\partial Q_n(\varphi)/\partial\varphi$). By Assumption A2, $\sqrt{n}\partial Q_n(\varphi_0)/\partial\varphi = O_p(1)$ *so that (because $h \to 0$)*

$$\left[\partial^2 Q_n(\bar{\varphi})/\partial\theta\partial\theta'\right]\sqrt{nh^{k/2}}(\hat{\theta} - \theta_0) = \left[\partial^2 Q_n(\bar{\varphi})/\partial\theta\partial\gamma'\right]\delta + o_p(1)$$

implying that $\sqrt{nh^{k/2}}(\hat{\theta} - \theta_0) = O_p(1)$ and $\hat{\theta} \xrightarrow{p} \theta_0$, since $\partial^2 Q_n(\bar{\varphi})/\partial\varphi\varphi'$ is $O_p(1)$, by Assumption A3. *Indeed,* Assumption A3 *implies $\partial^2 Q_n(\bar{\varphi})/\partial\varphi\partial\varphi' - J(\varphi_0) = o_p(1)$, and hence*

$$\sqrt{nh^{k/2}}(\hat{\theta} - \theta_0) = J_{\theta\theta}^{-1} J_{\theta\gamma}\delta + o_p(1), \tag{15}$$

so that

$$\begin{aligned} \sqrt{nh^{k/2}}\,(\hat{\varphi} - \varphi_0) &= \begin{pmatrix} J_{\theta\theta}^{-1} J_{\theta\gamma}\delta \\ -\delta \end{pmatrix} + o_p(1) \\ &= \xi + o_p(1). \end{aligned} \tag{16}$$

■

Remark 2 *$J_{\theta\gamma}\delta = o(1)$, for all $\delta \neq 0$, if and only if $J_{\theta\gamma} = o(1)$, in which case (of course) $\sqrt{n}(\hat{\theta} - \theta_0) = -J_{\theta\theta}^{-1}\sqrt{n}\partial Q_n(\varphi_0)/\partial\theta + o_p(1) = O_p(1)$.*

The following proves Theorem 1; a more detailed derivation can be found at:
http://personalpages.manchester.ac.uk/staff/chris.orme/

Proof of Theorem 1. Define $\hat{\varepsilon}_{ir} \equiv \varepsilon_r(W_i;\hat{\varphi})$, $\varepsilon_{ir}^0 \equiv \varepsilon_r(W_i;\varphi_0)$, $g_{ir}(\varphi) \equiv g_r(W_i;\varphi)$, $G_{ir}(\varphi) \equiv G_r(W_i;\varphi)$, $F_{ir}(\varphi) \equiv F_r(W_i;\varphi)$ and let $T_n(\varphi)$ have typical element $T_{nr}(\varphi)$, so that

$$\begin{aligned} nh^{k/2}T_{nr}(\hat{\varphi}) &= nh^{k/2}T_{nr}(\varphi_0) \\ &\quad + 2\frac{nh^{k/2}}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij}(\hat{\varepsilon}_{jr} - \varepsilon_{jr}^0) + \frac{nh^{k/2}}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{\varepsilon}_{ir} - \varepsilon_{ir}^0)K_{ij}(\hat{\varepsilon}_{jr} - \varepsilon_{jr}^0). \\ &= nh^{k/2}T_{nr}(\varphi_0) + 2T_{1nr} + T_{2nr}, \quad \text{say.} \end{aligned}$$

It suffices to show that $T_{1nr} = o_p(1)$, whilst $T_{2nr} = E\left[|d_r(X;\varphi_0)'\xi|^2 f(X)\right] + o_p(1)$:

1. T_{1nr} : write

$$\begin{aligned} \hat{\varepsilon}_{jr} - \varepsilon_{jr}^0 &= g_{jr}(\varphi_0)'(\hat{\varphi} - \varphi_0) + \frac{1}{2}(\hat{\varphi} - \varphi_0)'G_{jr}(\varphi_0)(\hat{\varphi} - \varphi_0) \\ &\quad + \frac{1}{6}vec((\hat{\varphi} - \varphi_0)(\hat{\varphi} - \varphi_0)')'F_{jr}(\bar{\varphi}^{(r)})(\hat{\varphi} - \varphi_0) \end{aligned}$$

where $\bar{\varphi}^{(r)}$ is a "mean value" such that $\left\|\bar{\varphi}^{(r)} - \varphi_0\right\| \leq \|\hat{\varphi} - \varphi_0\| = O_p(n^{-1/2}h^{-k/4})$. Substituting this expression into T_{1nr} yields

$$
\begin{aligned}
T_{1nr} &= h^{k/4}\left\{\frac{\sqrt{n}}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} g_{jr}(\varphi_0)'\right\}\xi_n \\
&\quad +\frac{1}{2}\xi_n'\left\{\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} G_{jr}(\varphi_0)\right\}\xi_n \\
&\quad +\frac{1}{6}vec(\xi_n\xi_n')'\left\{\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} F_{jr}(\bar{\varphi}^{(r)})\right\}(\hat{\varphi}-\varphi_0),
\end{aligned}
$$

where $\xi_n = \sqrt{nh^{k/2}}(\hat{\varphi} - \varphi_0) = \xi + o_p(1) = O_p(1)$. First, under Assumptions B1-4b and C2(i)&(ii), the results from Powell, Stock and Stoker (1989, Lemma 3.1) and Zheng (1996, Lemma 3.3b) imply that
$\frac{\sqrt{n}}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} g_{jr}(\varphi_0) = O_p(1)$. Second, Assumptions B1-4b , C3(i), and Powell, Stock and Stoker (1989, Lemma 3.1) imply that $\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} G_{jr}(\varphi_0) = o_p(1)$. Finally, Assumptions B1-3, C1 and C4 imply

$$
\begin{aligned}
E\left[\sup_{\varphi}\left\|\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} F_{jr}(\varphi)\right\|\right] &\leq \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}E[h^{-k}\left|\varepsilon_{ir}^0\right| K_{ij}P(W_j)] \\
&= O(1)
\end{aligned}
$$

so that $\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}\varepsilon_{ir}^0 K_{ij} F_{jr}(\bar{\varphi}^{(r)}) = O_p(1)$ by Markov's Inequality.
Thus, $T_{1nr} = o_p(1)$.

2. T_{2nr} : write

$$\hat{\varepsilon}_{ir} - \varepsilon_{ir}^0 = g_{ir}(\varphi_0)'(\hat{\varphi} - \varphi_0) + \frac{1}{2}(\hat{\varphi} - \varphi_0)'G_{ir}(\tilde{\varphi}^{(r)})(\hat{\varphi} - \varphi_0)$$

where $\tilde{\varphi}^{(r)}$ is a "mean value" such that $\left\|\tilde{\varphi}^{(r)} - \varphi_0\right\| \leq \|\hat{\varphi} - \varphi_0\| = O_p(n^{-1/2}h^{-k/4})$. Substituting this expression into T_{2nr} yields

$$T_{2nr} = \xi_n'\left\{\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}g_{ir}(\varphi_0)K_{ij}g_{jr}(\varphi_0)'\right\}\xi_n + R_n, \quad \text{say}$$

First, Assumptions B1-3, C2(ii)&(iv), and Powell, Stock and Stoker (1989, Lemma 3.1) imply that

$$\xi_n'\left\{\frac{1}{n(n-1)h^k}\sum_{i=1}^{n}\sum_{j\neq i}g_{ir}(\varphi_0)K_{ij}g_{jr}(\varphi_0)'\right\}\xi_n = E\left[\left|d_r(X;\varphi_0)'\xi\right|^2 f(X)\right] + o_p(1),$$

with $E\left[\left|d_r(X;\varphi_0)\xi\right|^2 f(X)\right] = O(1)$, and $d_r(X;\varphi) = E[\partial\varepsilon_r(W;\varphi)/d\varphi'|x]$, $r = 1, ..., m$. Second, Assumptions B1-3, C2(iii),C3(ii), $(\hat{\varphi} - \varphi_0) = O_p(n^{-1.2}h^{-k/4})$ and Markov's Inequality ensure that $R_n = o_p(1)$.
Thus, $T_{2nr} = E\left[\left|d_r(X;\varphi_0)'\xi\right|^2 f(X)\right] + o_p(1)$.

The assumptions ensure that $\mu(\varphi_0) = E\left[\|d(X;\varphi_0)\xi\|^2 f(X)\right] = O(1)$, and continuous in φ_0, so that $\lim_{n\to\infty}\mu(\varphi_0) = \mu_0$ exists. By (2), (ii) $nh^{k/2}T_n(\varphi_0)'\iota/\sqrt{\iota'\Omega_0\iota} \xrightarrow{d} N(0,1)$ and due to the local alternatives $\lim_{n\to\infty}\iota'\Omega_0\iota = \Sigma_0$. The result then follows, noting that: (i) $\sum_{r=1}^{m}\left|d_r(X;\varphi_0)'\xi\right|^2 = \|d(X;\varphi_0)'\xi\|^2$, and (ii) the dominance conditions which ensure that ensure that $\hat{\Sigma}_n$ is consistent under the null, also ensure that $\hat{\Sigma}_n - \Sigma_0 \xrightarrow{p} 0$ under the local alternatives so that $\mathcal{J}_n \xrightarrow{d} N(\Sigma_0^{-1/2}\mu_0, 1)$. ■

Table 1: Rejection Frequencies, DGP_0

	Nominal Significance Levels								
	$n=50$			$n=100$			$n=200$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
NORMAL ERRORS									
Normal Regressors									
$\mathcal{F}_n$	10.82	5.20	1.31	10.91	5.94	1.57	9.50	5.04	1.19
$\mathcal{H}_n$	10.10	4.86	1.03	9.93	5.00	1.29	10.26	5.29	1.23
$\mathcal{H}_n^K$	9.90	5.04	1.09	9.95	5.05	1.21	10.20	5.36	1.43
$\mathcal{S}_n$	10.62	5.51	1.39	10.99	5.80	1.34	10.16	5.25	1.09
$\mathcal{S}_n^M$	10.51	5.66	1.61	10.63	5.57	1.39	10.29	5.52	1.24
Chi-Square Regressors									
$\mathcal{F}_n$	10.89	5.56	1.42	10.28	5.25	1.36	10.21	5.17	1.34
$\mathcal{H}_n$	10.11	5.12	1.29	10.50	5.61	1.42	9.98	4.95	1.02
$\mathcal{H}_n^K$	9.65	4.81	1.22	10.42	5.37	1.19	9.99	4.91	1.10
$\mathcal{S}_n$	11.01	5.74	1.53	9.71	5.25	1.24	10.62	5.47	1.22
$\mathcal{S}_n^M$	10.29	5.32	1.43	10.22	5.64	1.48	10.25	5.36	1.19
CHI-SQUARE ERRORS									
Normal Regressors									
$\mathcal{F}_n$	9.31	4.78	1.24	8.86	4.31	0.99	9.04	4.31	0.64
$\mathcal{H}_n$	11.07	6.30	1.82	11.15	5.70	1.45	9.87	5.01	1.25
$\mathcal{H}_n^K$	8.35	4.30	0.94	9.16	4.59	1.24	9.30	4.70	1.04
$\mathcal{S}_n$	18.48	8.97	1.85	37.40	21.72	5.53	76.03	59.92	27.43
$\mathcal{S}_n^M$	55.74	42.70	21.21	79.03	68.80	46.70	96.09	92.80	81.04
Chi-Square Regressors									
$\mathcal{F}_n$	9.90	5.02	1.06	9.20	4.44	0.79	8.98	4.36	0.71
$\mathcal{H}_n$	11.18	5.91	1.74	10.35	5.73	1.59	9.68	4.86	1.29
$\mathcal{H}_n^K$	8.62	4.44	0.99	9.32	4.56	0.99	9.33	4.55	1.04
$\mathcal{S}_n$	36.33	22.20	7.90	69.16	51.25	22.13	96.40	91.59	70.36
$\mathcal{S}_n^M$	67.06	54.16	30.44	90.56	84.11	64.79	99.10	98.07	93.89

Table 2: Rejection Frequencies, 5% Significance Level, Homoskedasticity

	DGP_1			DGP_2			DGP_3		
$n=$	50	100	200	50	100	200	50	100	200
NORMAL ERRORS									
Normal Regressors									
$\mathcal{F}_n$	19.78	37.86	70.22	12.56	19.14	38.46	28.16	53.34	88.50
$\mathcal{H}_n$	12.48	18.72	31.50	5.84	5.70	6.50	4.88	5.24	5.74
$\mathcal{H}_n^K$	4.20	5.18	11.04	4.70	5.30	5.96	4.56	4.58	5.30
$\mathcal{S}_n$	20.28	35.30	58.92	12.44	16.98	29.42	26.44	42.32	68.28
$\mathcal{S}_n^M$	7.60	7.60	9.60	6.30	6.10	5.06	7.38	7.52	6.90
Chi-Square Regressors									
$\mathcal{F}_n$	37.62	66.64	92.94	15.70	26.28	48.70	62.02	92.98	99.90
$\mathcal{H}_n$	13.08	20.44	35.40	4.48	4.22	4.66	4.02	3.28	4.20
$\mathcal{H}_n^K$	5.14	4.60	7.02	4.90	5.02	5.24	3.74	3.54	4.56
$\mathcal{S}_n$	32.68	55.08	80.80	13.38	20.76	34.12	50.46	77.94	96.00
$\mathcal{S}_n^M$	12.46	16.92	23.42	5.48	5.96	6.10	11.22	11.78	13.62
CHI-SQUARE ERRORS									
Normal Regressors									
$\mathcal{F}_n$	23.90	41.84	72.00	15.24	22.98	41.26	32.20	60.32	90.56
$\mathcal{H}_n$	16.82	20.20	24.88	8.26	6.74	7.14	8.52	7.34	7.00
$\mathcal{H}_n^K$	3.90	5.18	6.84	4.30	4.20	4.32	5.08	5.54	6.14
$\mathcal{S}_n$	13.42	22.56	46.98	12.34	22.22	46.66	37.60	61.46	86.30
$\mathcal{S}_n^M$	51.00	71.60	89.18	45.70	69.80	91.30	21.34	32.52	56.22
Chi-Square Regressors									
$\mathcal{F}_n$	42.56	69.66	93.18	16.82	30.02	49.94	65.44	91.84	99.70
$\mathcal{H}_n$	12.90	17.46	21.50	5.98	4.66	5.00	8.00	7.14	7.14
$\mathcal{H}_n^K$	4.78	6.08	6.44	4.30	4.10	4.84	6.12	6.28	6.04
$\mathcal{S}_n$	51.76	77.30	94.22	22.22	48.38	83.72	42.88	59.70	77.62
$\mathcal{S}_n^M$	35.52	47.28	63.92	51.90	77.26	94.48	47.22	67.86	86.48

Table 2a: Increase in Rejection Frequencies, 5% Significance Level, Homoskedasticity
Sensitivity to Misspecified Regression Functional Form

	DGP_1			DGP_2			DGP_3		
$n =$	50	100	200	50	100	200	50	100	200
NORMAL ERRORS									
Normal Regressors									
$\mathcal{F}_n$	15	32	65	7	13	33	23	48	83
$\mathcal{H}_n$	8	14	26	1	1	1	0	0	0
$\mathcal{H}_n^K$	-1	0	6	0	0	1	0	0	0
$\mathcal{S}_n$	15	30	54	7	11	24	21	37	63
$\mathcal{S}_n^M$	2	2	4	1	1	0	2	2	1
Chi-Square Regressors									
$\mathcal{F}_n$	32	61	88	10	21	44	56	88	95
$\mathcal{H}_n$	8	15	30	-1	-1	0	-1	-2	0
$\mathcal{H}_n^K$	0	-1	2	0	0	0	-1	-2	0
$\mathcal{S}_n$	27	50	75	8	16	29	45	73	91
$\mathcal{S}_n^M$	7	11	18	0	0	1	6	6	8
CHI-SQUARE ERRORS									
Normal Regressors									
$\mathcal{F}_n$	19	38	68	10	19	37	27	56	86
$\mathcal{H}_n$	11	14	20	2	1	2	2	2	2
$\mathcal{H}_n^K$	0	1	2	0	0	0	1	1	1
$\mathcal{S}_n$	4	1	-13	3	1	-13	29	40	26
$\mathcal{S}_n^M$	8	3	-4	3	1	-2	-21	-36	-37
Chi-Square Regressors									
$\mathcal{F}_n$	38	65	89	12	26	46	60	87	95
$\mathcal{H}_n$	7	12	17	0	-1	0	2	1	2
$\mathcal{H}_n^K$	0	2	2	0	0	0	2	2	1
$\mathcal{S}_n$	30	26	3	0	-3	-8	21	8	-14
$\mathcal{S}_n^M$	-19	-37	-34	-2	-7	-4	-7	-16	-12

Table 2b: Increase in Rejection Frequencies, 5% Significance Level, Homoskedasticity
Sensitivity to Asymmetric Errors

	DGP_0			DGP_1			DGP_2			DGP_3		
$n =$	50	100	200	50	100	200	50	100	200	50	100	200
					Normal Regressors							
$\mathcal{F}_n$	0	-2	-1	4	4	2	3	4	3	4	6	2
$\mathcal{H}_n$	1	1	0	4	1	-7	2	1	1	4	2	1
$\mathcal{H}_n^K$	-1	0	-1	0	0	-4	0	-1	-2	1	1	1
$\mathcal{S}_n$	3	16	55	-7	-13	-12	0	5	17	11	19	18
$\mathcal{S}_n^M$	37	63	87	43	64	80	39	64	86	14	25	49
					Chi-Square Regressors							
$\mathcal{F}_n$	-1	-1	-1	5	3	1	1	4	1	3	-1	0
$\mathcal{H}_n$	1	0	0	0	-3	-14	2	0	0	4	4	3
$\mathcal{H}_n^K$	0	-1	0	0	1	-1	-1	-1	0	2	3	1
$\mathcal{S}_n$	16	46	86	19	22	13	9	28	50	-8	-18	-18
$\mathcal{S}_n^M$	49	78	93	23	30	41	46	71	88	36	56	73

Table 3: Rejection Frequencies, 5% Significance Level, Heteroskedasticity

	DGP_4			DGP_5			DGP_6			DGP_7		
$n =$	50	100	200	50	100	200	50	100	200	50	100	200
NORMAL ERRORS												
Normal Regressors												
$\mathcal{F}_n$	6.42	5.00	4.64	13.46	22.16	41.58	10.62	14.16	23.76	19.30	35.08	62.30
$\mathcal{H}_n$	27.96	50.08	80.18	34.42	63.72	91.76	26.80	49.70	81.16	25.68	45.36	76.00
$\mathcal{H}_n^K$	10.20	20.98	49.24	9.24	22.08	56.66	9.64	19.74	50.48	8.46	18.32	43.22
$\mathcal{S}_n$	7.32	4.88	5.42	12.14	19.00	29.64	9.24	10.66	13.34	15.40	20.86	28.42
$\mathcal{S}_n^M$	5.94	4.80	5.50	7.92	9.78	11.54	7.46	7.24	7.80	10.02	12.20	17.54
Chi-Square Regressors												
$\mathcal{F}_n$	7.90	6.92	6.90	27.54	47.24	73.76	13.20	19.30	34.14	44.30	72.16	95.12
$\mathcal{H}_n$	28.04	48.90	77.16	33.00	59.40	87.52	27.10	47.76	74.50	21.94	39.76	67.12
$\mathcal{H}_n^K$	10.12	17.14	40.48	9.74	19.16	40.32	9.12	17.56	36.46	7.50	12.54	28.00
$\mathcal{S}_n$	7.08	7.20	6.28	23.46	34.50	48.64	12.32	14.04	18.30	34.12	48.54	63.56
$\mathcal{S}_n^M$	6.00	5.66	5.24	12.56	16.06	22.36	7.94	8.10	9.34	15.36	21.08	31.58
CHI-SQUARE ERRORS												
Normal Regressors												
$\mathcal{F}_n$	5.30	4.28	4.04	14.94	23.46	42.80	10.28	14.80	23.82	23.00	38.52	65.34
$\mathcal{H}_n$	22.78	32.82	46.78	26.58	38.48	59.84	23.34	32.44	46.36	24.92	33.86	48.10
$\mathcal{H}_n^K$	5.66	7.66	11.92	5.18	7.44	13.26	4.84	7.16	10.42	7.30	9.60	15.90
$\mathcal{S}_n$	9.82	16.50	37.18	11.66	21.58	51.20	9.78	18.18	41.90	22.76	31.74	46.16
$\mathcal{S}_n^M$	32.30	54.60	81.80	42.04	58.18	79.30	40.44	60.82	84.38	23.80	38.80	62.58
Chi-Square Regressors												
$\mathcal{F}_n$	6.70	7.46	6.50	29.58	50.28	79.04	13.54	18.92	33.48	50.36	73.68	92.92
$\mathcal{H}_n$	22.46	29.18	43.26	24.92	35.92	53.04	20.22	29.32	40.96	23.18	33.28	47.00
$\mathcal{H}_n^K$	7.18	8.54	12.68	7.14	9.30	14.36	5.36	7.46	11.46	7.24	9.32	13.56
$\mathcal{S}_n$	19.10	37.66	66.36	43.86	65.16	82.40	21.20	41.02	70.96	30.54	36.90	49.66
$\mathcal{S}_n^M$	44.42	68.54	86.34	35.78	46.00	58.28	46.28	66.76	84.58	49.48	70.46	88.34

Table 3a: Increase in Rejection Frequencies, 5% Significance Level, Heteroskedasticity Sensitivity to Misspecified Regression Functional Form

	DGP_5			DGP_6			DGP_7		
$n =$	50	100	200	50	100	200	50	100	200
NORMAL ERRORS									
Normal Regressors									
$\mathcal{F}_n$	7	17	37	4	9	19	13	30	58
$\mathcal{H}_n$	6	14	12	-1	0	1	-2	-5	-4
$\mathcal{H}_n^K$	-1	1	7	-1	-1	1	-2	-3	-6
$\mathcal{S}_n$	5	14	24	2	6	8	8	16	23
$\mathcal{S}_n^M$	2	5	6	2	2	2	4	7	12
Chi-Square Regressors									
$\mathcal{F}_n$	20	40	67	5	12	27	36	65	88
$\mathcal{H}_n$	5	11	10	-1	-1	-3	-6	-9	-10
$\mathcal{H}_n^K$	0	2	0	-1	0	-4	-3	-5	-12
$\mathcal{S}_n$	16	27	42	5	7	12	27	41	57
$\mathcal{S}_n^M$	7	10	17	2	2	4	9	15	26
CHI-SQUARE ERRORS									
Normal Regressors									
$\mathcal{F}_n$	10	19	39	5	11	20	18	34	61
$\mathcal{H}_n$	4	16	13	1	0	0	2	1	1
$\mathcal{H}_n^K$	0	0	1	-1	-1	-2	2	2	4
$\mathcal{S}_n$	2	5	14	0	2	5	13	15	9
$\mathcal{S}_n^M$	10	4	-3	8	6	3	-9	-16	-19
Chi-Square Regressors									
$\mathcal{F}_n$	23	43	73	7	11	27	44	66	86
$\mathcal{H}_n$	2	7	10	-2	0	-2	1	4	4
$\mathcal{H}_n^K$	0	1	2	-2	-1	-1	0	1	1
$\mathcal{S}_n$	25	28	16	2	3	5	11	-1	-17
$\mathcal{S}_n^M$	-9	-23	-28	2	-2	-2	5	2	2

Table 3b: Increase in Rejection Frequencies, 5% Significance Level, Heteroskedasticity
Sensitivity to Asymmetric Errors

	DGP_4			DGP_5			DGP_6			DGP_7		
$n =$	50	100	200	50	100	200	50	100	200	50	100	200
						Normal Regressors						
$\mathcal{F}_n$	-1	-1	-1	1	1	1	0	1	0	4	3	3
$\mathcal{H}_n$	-5	-17	-33	-8	-25	-32	-3	-17	-35	-1	-12	-28
$\mathcal{H}_n^K$	-5	-13	-37	-4	-15	-43	-5	-13	-40	-1	-9	-27
$\mathcal{S}_n$	3	12	32	0	3	22	1	8	29	7	11	18
$\mathcal{S}_n^M$	26	50	76	34	48	68	33	54	77	14	27	45
						Chi-Square Regressors						
$\mathcal{F}_n$	-1	1	0	2	3	5	0	0	-1	6	2	-2
$\mathcal{H}_n$	-6	-20	-34	-8	-23	-34	-7	-18	-34	1	-6	-20
$\mathcal{H}_n^K$	-3	-9	-28	-3	-10	-26	-4	-10	-25	0	-3	-14
$\mathcal{S}_n$	12	30	60	20	31	24	9	27	53	-4	-12	-14
$\mathcal{S}_n^M$	38	63	81	23	30	36	38	59	75	34	49	57