Consistent Model Specification Testing

James Davidson and Andreea G. Halunga
Department of Economics
University of Exeter Business School
Rennes Drive
Exeter EX4 4PU
U.K.

October 2010

Abstract

We generalize Bierens' (1982, 1990) approach to a wider class of models and estimators. Bierens constructs consistent moment tests in the context of linear and nonlinear least squares but there are a number of mis-specifications, such as heteroskedastic errors, against which they will not typically have power. Our framework is independent of the form of the model, and covers all variants of maximum likelihood and quasi-maximum likelihood estimation and also the generalized method of moments. It has particular applications in new cases such as discrete data models, but the chief appeal of our approach is that it provides a "one size fits all" test. We specify a test based on a linear combination of individual components of the indicator vector that can be computed routinely, does not need to be tailored to the particular model, and is expected to have power against a wide class of alternatives. Although primarily envisaged as a test of functional form, this type of moment test can also be extended to testing for omitted variables.

JEL classification: C12

Keywords: specification testing; quasi-maximum likelihood estimators; generalized method of moments estimators

Corresponding Author:
Andreea G. Halunga,
Department of Economics,
University of Exeter Business School,
Streatham Court, Rennes Drive, Exeter EX4 4PU, UK.
Email: a.g.halunga@exeter.ac.uk

1 Introduction

Specification testing of econometric models frequently faces the difficulty that the investigator does not know what type of specification error to look for. Tests of functional form need to have power against a bewildering variety of possible alternatives. To compute tests of the Lagrange multiplier and Durbin-Hausman-Wu types one needs to specify, and in the latter case to estimate, a dummy alternative hypothesis. Authors such as Davidson and MacKinnon (1981) have documented how test power can depend critically on the alternative chosen to construct the test. There are always some alternatives against which a test will lack even consistency.

A general class of tests, of which most specification tests can be constructed as special cases, are the conditional moment tests of Newey (1985) and Tauchen (1985). The sample mean of a function depending on data and parameters is constructed, of which the population mean is zero under the null hypothesis. Typically, in applications the quantity in question is the product of model residuals, or normalized squared residuals, with a test indicator function (weighting function) depending on conditioning variables. Even though they are not typically constructed with a specific alternative in mind, these tests are generally not ‘consistent’, in the sense of rejecting the null hypothesis in a large enough sample against any deviation from the null model. Their power against specific alternatives depends on the choice of the weighting functions. However, Bierens (1982,1990) has suggested consistent model specification tests. By the use of an exponential weighting function these statistics in effect test an infinite set of moment conditions, in the context of linear or nonlinear least squares estimation. In the time series case, generalizations have been proposed by Bierens (1984, 1987), de Jong (1996) and Bierens and Ploberger (1997) with the latter generalizing a version of the integrated conditional moment test of Bierens (1982). Furthermore, Whang (2000, 2001) and Delgado, Dominguez and Lavergne (2006) propose consistent tests in an i.i.d. context by using an indicator function instead of the exponential weighting function of Bierens. The former tests are generalizations of both the Kolmogorov-Smirnov and Cramer-von-Mises statistics, whereas the latter authors consider only the Cramer-von Mises type test. Escanciano (2007) provides a unified theory for both continuous and discontinuous weighting functions using residual marked empirical processes in order to detect misspecifications in time series regression models. In semiparametric dynamic models, Chen and Fan (1999) extend the Bierens (1990) approach to testing conditional moment restrictions using the weighted integrated squared metric.

Another approach for constructing consistent tests of functional form is by comparing the fitted parametric regression function with a nonparametric model. Some examples of such tests for i.i.d. data have been proposed by Zheng (1996), Eubank and Spiegelman (1990), Härdle and Mammen (1993), Hong and White (1996), Fan and Li (1996a), inter alia, whereas for time series developments include Fan and Li (1996b), Koul and Stute (1999) and Dominguez and Lobato (2003). Although these tests are consistent against all alternatives to the null hypothesis, they have nontrivial power only under the local alternatives that approach the null at a rate slower than $T^{-1/2}$ which decreases as well due to the curse of dimensionality, where T is the sample size. Further, such tests depend on a smoothing parameter whose choice is not trivial and this will influence the results.

The idea that we develop in this paper is to generalize Bierens’ approach to a much wider class of models and estimators. Our framework extends to cover all variants of maximum likelihood and quasi-maximum likelihood estimation and also the generalized method of moments. Parameter estimation is done in these cases by solving the equations obtained by equating to zero a set of functions of data and parameters, which we refer to generically as the scores. In a sample of size T these functions consist of sums of T terms, the ‘score contributions’, that sum to zero by construction at the estimated point. The rationale for the choice of the estimator, in each case, is that under the hypotheses of the model the score contributions evaluated at the 'true' parameter values have individual means of zero, conditional on a designated set of conditioning variables, with probability 1.[1] Here, 'true' may mean that economic theory assigns a specific interpretation to the parameter values, or simply that these are the values that solve the respective equations when our maintained hypotheses hold. In the latter case it may be strictly more correct to speak of an 'adequate' model specification than a correct one, and under this interpretation we may sometimes prefer to call these the 'pseudo-true' values. The minimal requirement, trivial with i.i.d. data, is that the same set of values characterize each observation in the sample.

In either case, our object is consistent estimation of the parameters satisfying the condition. Our maintained hypotheses typically include a list of included variables and a functional form and, most importantly, the designation of the variables that can be validly treated as fixed in forming conditional expectations, which we henceforth refer to as 'exogenous'. This exogeneity property is related to, though not identical with, the weak exogeneity condition defined by Engle, Hendry and Richard (1983). Note that it depends on the interpretation of the model and is not a condition subject to verification in the data.

Under correct specification, so defined, it follows that functions of the exogenous variables should be uncorrelated with the score contributions. Since the scores are frequently represented as the sum of products of a residual and another function of the data, where correct specification imposes a special property on the residual, we often think of these moment tests as tests of the residuals. However, there are important classes of models for which a residual may not be well, or uniquely, defined. For these cases, specification tests can be developed directly in terms of score contributions. This framework may be particularly useful for developing consistent specification tests in particular cases, such as discrete choice models, but the chief appeal of our approach is the "one size fits all" principle. We specify a test based on a linear combination of individual components of the indicator vector, that can be computed routinely and does not need to be tailored to each particular model, and yet will have power against a wide class of alternatives. Although primarily envisaged as a test of functional form, this type of moment test can be extended to testing for omitted variables by defining the weighting functions appropriately. The present work focuses on the case of independently distributed observations. A companion paper will consider the extension to tests of dynamic specification.

The paper is organized as follows, Section 2 briefly recounts the Bierens framework for constructing a consistent test of functional form of the mean equation. Section 3 develops a consistent specification test based on the score approach. In Section 4.1, we present Monte Carlo evidence on a variety of likelihood-based applications including nonlinear and heteroskedastic regressions, discrete choice and count data models. Section 4.2 then looks at the application to models with simultaneity, and instrumental variables estimation, and Section 5 concludes the paper. All proofs, together with some supporting lemmas, are collected in the Appendix.

2 The Bierens' framework

Consider the following model indexed by $\beta \in \Theta \subset R^k$

$$y_t = m(x_t, \beta) + \varepsilon_t \tag{2.1}$$

where for each $\beta \in \Theta$, $m(\cdot)$ is a known Borel measurable real valued function of the $(k \times 1)$ vector of regressors x_t. $\{(x_t', \varepsilon_t)\}$ is a sequence of *i.i.d.* random vectors. The null hypothesis of

[1] This statement may need qualifying in the GMM case, as we explain in Section 5 below.

correct specification of the model in (2.1) is

$$P\left(E\left[y_t|x_t\right] = m\left(x_t, \beta_0\right)\right) = 1 \text{ for some } \beta_0 \in \Theta \tag{2.2}$$

whereas the alternative hypothesis is

$$P\left(E\left[y_t|x_t\right] = m\left(x_t, \beta\right)\right) < 1 \text{ for all } \beta \in \Theta.$$

Tests of the null hypothesis in (2.2) have been constructed based on the correlation of the least squares residuals with a function of the data. These are often characterized as conditional moment tests, as in Newey (1985) and Tauchen (1985).

Let the unknown parameter β_0 be estimated by the least squares estimator

$$\hat{\beta} = \arg\min_{\beta\in\Theta} T^{-1} \sum_{t=1}^{T} \left(y_t - m\left(x_t, \beta\right)\right)^2.$$

The Bierens' (1990) consistent specification test of functional form is based on the proximity to zero of the indicator function

$$r_T(\hat{\beta}, \xi) = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t w\left(x_t, \xi\right), \tag{2.3}$$

where $\hat{\varepsilon}_t = y_t - m(x_t, \hat{\beta})$ is the residual process and $w_t = w\left(x_t, \xi\right)$ is a nonlinear transformation of the regressors where $\xi \in \Xi$ with Ξ a compact subset of $\mathbb{R}^k$.

Conditioning on x_t is equivalent to conditioning on any Borel-measurable isomorphic function of x_t (see, e.g., Davidson, 1994, Theorem 10.3) such that the function will generate the same σ-field as x_t. Bierens (1982, 1990) suggests using the weight function

$$w\left(x_t, \xi\right) = \prod_{i=1}^{k} \exp\left(\xi_i \varphi\left(\tilde{x}_{ti}\right)\right) \tag{2.4}$$

where φ is a one-to-one mapping from $\mathbb{R}$ to $\mathbb{R}$ chosen by Bierens (1990) as $\varphi\left(\tilde{x}_{ti}\right) = \arctan \tilde{x}_{ti}$, for $i = 1, .., k$ where $\tilde{x}_{ti}$ are in the standardized form

$$\tilde{x}_{ti} = \frac{x_{ti} - \bar{x}_i}{s_i}$$

with $\bar{x}_i$ and s_i representing the sample mean and sample standard deviation of x_{ti}, respectively, to avoid the problem of the weight function being invariant due to scale factors. The choice of the exponential in the weight function (2.4) is not crucial. As shown by Stinchcombe and White (1998) any function that admits an infinite series approximation on compact sets, with non-zero series coefficients, could in principle be employed to construct a consistent test.

Under the null hypothesis and standard regularity conditions, taking a mean value expansion of $\sqrt{T} r_T(\hat{\beta}, \xi)$ around β_0, Bierens (1982, 1990) shows that

$$\sqrt{T} r_T(\hat{\beta}, \xi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{0t} \phi_t\left(\xi\right) + o_p\left(1\right)$$

uniformly over Ξ, where $\varepsilon_{0t} = y_t - m\left(x_t; \beta_0\right)$ and

$$\phi_t\left(\xi\right) = w\left(x_t, \xi\right) - b\left(\xi\right)' A^{-1} \left.\frac{\partial m\left(x_t, \beta\right)}{\partial \beta'}\right|_{\beta=\beta_0}$$
$$b\left(\xi\right) = E\left[\frac{\partial m\left(x_t, \beta\right)}{\partial \beta'} w\left(x_t, \xi\right)\right]_{\beta=\beta_0}$$

$$A = E\left[\frac{\partial m\left(x_t,\beta\right)}{\partial\beta}\frac{\partial m\left(x_t,\beta\right)}{\partial\beta'}\right]_{\beta=\beta_0}.$$

Bierens (1990) shows that under the null hypothesis (2.2), $\sqrt{T}r_T(\hat{\beta},\xi)$ converges weakly in distribution to a zero-mean Gaussian process on Ξ, with covariance function

$$\Pi\left(\xi_1,\xi_2\right) = E[\varepsilon_t^2\phi_t\left(\xi_1\right)\phi\left(\xi_2\right)]_{\beta=\beta_0}. \tag{2.5}$$

The Bierens test based on the sample moment (2.3) with the weight function in (2.4) entails an infinite number of moment conditions. The consistency of the test is based on the result shown in Lemma 1 of Bierens (1990), that under the alternative hypothesis and specified regularity conditions,

$$E\left[(y_t - m(x_t,\beta_0))w\left(x_t,\xi\right)\right] = 0 \tag{2.6}$$

only when ξ belongs to a set of Lebesgue measure zero. Consistency can therefore be achieved with a suitable choice of ξ. If ξ is chosen from a continuous distribution, the moment in (2.6) will be non-zero with probability one. The Bierens (1990) test statistic is constructed as

$$\hat{B} = \sup_{\xi\in\Xi} B\left(\xi\right) \tag{2.7}$$

where

$$B\left(\xi\right) = \frac{\left(\sqrt{T}r_T(\hat{\beta},\xi)\right)^2}{\hat{\Pi}^*\left(\xi\right)}$$

with $\hat{\Pi}^*\left(\xi\right)$ a consistent estimator of $\Pi^*\left(\xi\right) = \Pi\left(\xi,\xi\right)$ defined in (2.5). A similar statistic to the time series case have been developed by de Jong (1996) in which Ξ grows in dimension to infinity with the sample size. Another choice for the construction of a consistent test for functional form of the conditional mean equation is the Cramėr-von-Mises functional as suggested by Bierens (1982) and Bierens and Ploberger (1997)[2] .

The Bierens test is specifically designed for possible nonlinear models estimated by nonlinear least squares. However, this test can also be constructed in the conditional moment test framework, and QMLE applied to obtain a consistent estimator of β_0. Nevertheless, the consistent tests of Bierens (1982, 1990) and Bierens and Ploberger (1997) are not designed against misspecification in second moments, and are suitable only for models for which a properly defined residual is available. There are important cases, such as discrete choice models, where there is no unique generalization of a test based on residuals. However, specification tests for such models are often constructed based on a suitable defined score, and it is from this approach that we take our cue in the next section.

3 A consistent test based on the score contributions

Bierens' procedure is an application of the conditional moment test principle in which the property of the regression residuals of having zero conditional mean in the true model is exploited. There may be other functions available to which the same property can be attributed. However,

[2] Bierens (1982) constructs a consistent test as $\int_\Xi \hat{T}_B\left(\xi\right)d\xi$ but could not establish the type of the limiting distribution, but only its first moment. He derives upper bounds of the critical values based on the Chebyshev's inequality for first moments. Bierens and Ploberger (1997) obtain the limiting distribution of the integrated conditional moment test and since the critical values depend on the data generating process, they derive case-independent upper bounds of the critical values.

functions that exist for virtually any parametric model are the scores of the estimation criterion. In this section we propose a consistent specification test applicable to a very general class of continuous distributed models.

Consider independently sampled variables $(y_t', z_t')'$, where y_t $(G \times 1)$ is a vector of dependent variables and z_t $(K \times 1)$ is a vector of exogenous variables. Defining for $k \leq K$ a subvector x_t $(k \times 1)$ of z_t, where $x_t = z_t$ is possible, our parametric model can be taken as defined by a p-vector of functions $d_t(\theta) = d(y_t, x_t, \theta)$ for $\theta \in \Theta \subset \mathbb{R}^p$, such that there exists a vector of parameters of interest $\theta_0 \in \text{int}(\Theta)$ satisfying

$$E\left[d_t\left(\theta_0\right)|z_t\right] = 0 \text{ w.p.1} \tag{3.1}$$

In many cases we shall of course have that $d_t = \partial l_t / \partial \theta$ where $l_t(\theta)$ is a log-likelihood contribution, or similar, satisfying the condition that $E\left[l_t(\theta)|z_t\right]$ is maximized at θ_0 with probability 1, subject to regularity conditions ensuring that (3.1) holds. Given a sample of data indexed by $t = 1, \ldots, T$, we accordingly expect to estimate θ_0 consistently by

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L_T(\theta)$$

where $L_T(\theta) = \sum_{t=1}^{T} l_t(\theta)$ represents the appropriate sample criterion function. Accordingly the estimate $\hat{\theta}$ is constructed as a solution to

$$\frac{1}{T}\sum_{t=1}^{T} d_t(\hat{\theta}) = 0. \tag{3.2}$$

We shall refer to the components d_t generically as the score contributions, although note that θ_0 could be defined directly by an orthogonality condition in which case estimation would be done by the method of moments. We shall subsequently (see Section 4.2) consider models where d_t depends on the full sample and hence is strictly an array, and (3.1) is valid asymptotically but not necessarily for finite T.

Note that in this context, exogeneity of z_t is defined by the condition that θ_0 satisfies (3.1) almost surely, and in this sense it is a condition defined by the interpretation of the model. Our correct specification condition does not entail that the conditioning variables need to be used to construct the criterion. When $k < K$, condition (3.1) embodies the assumption of correct exclusion from the model of some valid conditioning variables. z_t may include any exogenous variable that may legitimately contribute to the explanation of y_t, and this set could in principle be very large, although our procedure puts limits on it in practice. This allows us to consider problems of omitted variables, although the case $x_t = z_t$ would apply in many cases where the specification issue relates solely to functional form.

Following the approach of Bierens (1982, 1990), the null hypothesis of correct specification might be stated in the form

$$P\left(E\left[d_t\left(\theta_0\right)|z_t\right] = 0\right) = 1 \text{ for } t = 1, .., T \tag{3.3}$$

with alternative hypothesis

$$P\left(E\left[d_t(\theta)|z_t\right] = 0\right) < 1, \text{ for all } \theta \in \Theta \text{ and at least one } t. \tag{3.4}$$

Be careful to note that we define a 'consistent test' in terms of rejections when (3.4) holds. However, we cannot entirely rule out the possibility that, even in certain cases that we might wish to call misspecification, (3.3) remains true. To take a leading example, in the continuous data case, the null hypothesis is generally satisfied even if the true distribution is not the one assumed for constructing the criterion function, as when the test is based on the quasi-likelihood function. In fact, we see this feature of our test as advantageous. It is arguably a desirable feature that deviations from the true distribution of the data are not detected as long as the estimators are consistent and asymptotically normal. A parallel case exists in the class of count data models to be discussed in Section 4.4. We find cases where the null hypothesis in (3.3) is true in spite of misspecification of the distribution as a whole, such that the Poisson distribution defines a quasi-maximum likelihood estimator; see Gourieroux, Montfort and Trognon (1984).

We test (3.3) by a conditional moment test on the covariance between these score contributions and a suitable measurable function of exogenous variables. Noting that $\hat{\theta}$ is defined by (3.2), our test indicator is

$$s_T(\hat{\theta}, \xi) = \frac{1}{T} \sum_{t=1}^{T} d_t(\hat{\theta}) w_t(\xi) \tag{3.5}$$

where $w_t(\xi) = w(z_t, \xi)$ is the weight function in (2.4) and $\xi \in \Xi$ with Ξ a compact subset of $\mathbb{R}^K$.

The following assumptions constitute the maintained hypothesis, in which context we derive our tests. Throughout this paper, $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix.

Assumptions

1. The observed data $(y_t', z_t')'$, $t = 1, .., T$, form a sequence of independently distributed random variables.

2. The parameter space Θ is a compact subspace of $\mathbb{R}^p$.

3. $d_t(\theta) : \mathbb{R}^{G+k} \times \Theta \longmapsto \mathbb{R}^p$ is a Borel measurable function for each $\theta \in \Theta$ and continuously differentiable on Θ.

4. For all t and some $s > 0$, the following are bounded uniformly in t,

   (i) $E\left[\sup_{\theta \in \Theta} \|d_t(\theta)\|^{2(1+s)}\right]$,

   (ii) $E\left[\sup_{\theta \in \Theta, \xi \in \Xi} \|d_t(\theta) w_t(\xi)\|^{2(1+s)}\right]$,

   (iii) $E\left[\sup_{\theta \in \Theta} \left\|\partial d_t(\theta)/\partial \theta'\right\|^{1+s}\right]$,

   (iv) $E\left[\sup_{\theta \in \Theta, \xi \in \Xi} \left\|\partial d_t(\theta)/\partial \theta' w_t(\xi)\right\|^{1+s}\right]$.

5. The matrix

   $$M = -\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\partial d_t(\theta)/\partial \theta'\right]_{\theta = \theta_0} \tag{3.6}$$

   is finite and non-singular;

6. Under the null hypothesis, $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, M^{-1} \Sigma M^{-1}\right)$, where M is defined in (3.6) and

   $$\Sigma = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\left[d_t(\theta) d_t(\theta)'\right]_{\theta = \theta_0} < \infty. \tag{3.7}$$

The following lemmas will provide the basis for the consistent test.

Lemma 3.1 *If $P(E[d_t(\theta)|z_t] = 0) < 1$, then the set*

$$B = \left\{\xi \in \mathbb{R}^K : \; E[d_t(\theta) w_t(\xi)] = 0\right\}$$

has Lebesgue measure zero for any $\theta \in \Theta$.

Lemma 3.2 *Under Assumptions 1-6 and H_0 in (3.3)*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} d_t(\hat{\theta})w_t(\xi) \xrightarrow{d} N\left(0, V(\xi)\right)$$

pointwise in the set of ξ, where

$$V(\xi) = R(\xi) - Q(\xi) M^{-1} P(\xi)' - P(\xi) M^{-1} Q(\xi)' + Q(\xi) M^{-1} \Sigma M^{-1} Q(\xi)' \tag{3.8}$$

and

$$Q(\xi) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[-w_t(\xi)\frac{\partial d_t(\theta)}{\partial\theta'}\right]_{\theta=\theta_0} \tag{3.9}$$

$$P(\xi) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[w_t(\xi)d_t(\theta)d_t(\theta)'\right]_{\theta=\theta_0} \tag{3.10}$$

$$R(\xi) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[w_t(\xi)^2 d_t(\theta)d_t(\theta)'\right]_{\theta=\theta_0}. \tag{3.11}$$

The covariance matrix $V(\xi)$ in (3.8) can be consistently estimated by

$$\hat{V}(\xi) = \hat{R}(\xi) - \hat{Q}(\xi) \hat{M}^{-1} \hat{P}(\xi)' - \hat{P}(\xi) \hat{M}^{-1} \hat{Q}(\xi)' + \hat{Q}(\xi) \hat{M}^{-1} \hat{\Sigma} \hat{M}^{-1} \hat{Q}(\xi)' \tag{3.12}$$

where hats denote evaluation at the consistent estimator $\hat{\theta}$.

Assumption 7 The set $B^* = \left\{\xi \in \mathbb{R}^K : \operatorname{rank}(V(\xi)) < p\right\}$ has Lebesgue measure zero.

Subject to Assumption 7, a joint consistent specification test can be constructed based on the test indicator $s_T(\hat{\theta}, \xi)$ defined in (3.5) that takes into account all the components of the score vector. This is as follows

$$S_B(\xi) = \frac{1}{T}\left(\sum_{t=1}^{T} d_t(\hat{\theta})w_t(\xi)\right)' \hat{V}(\xi)^{-1}\left(\sum_{t=1}^{T} d_t(\hat{\theta})w_t(\xi)\right). \tag{3.13}$$

Note that $V(\xi)$ should have rank p under the same circumstances that Σ has rank p for all ξ except on a set of Lebesgue measure zero. Provided that x_t is a linearly independent set of variables, the case $\xi = 0$ appears to be the unique counter-example under which we should obtain $V(\xi) = 0$.

The asymptotic distribution of the joint test statistic in (3.13) for each ξ is established in the following Theorem.

Theorem 3.1 *For every $\xi \in \mathbb{R}^K / B_0 \cup B^*$, where B_0 is the set defined in Lemma 3.1 for the case $\theta = \theta_0$, and B^* is the set defined in Assumption 7, the joint test $S_B(\xi)$ in (3.13) under H_0 in (3.3) has a limiting chi-square distribution with p degrees of freedom, whereas under H_1 in (3.4), $S_B(\xi)/T \to q(\xi)$ a.s., where $q(\xi) > 0$.*

One way to implement this test would be to choose the vector ξ arbitrarily, but following the approach of Bierens (1990) we anticipate the greatest power would be obtained by considering the statistic

$$\hat{S}_B = \sup_{\xi\in\Xi} S_B(\xi). \tag{3.14}$$

Where z_t is a vector, the choice of ξ will determine the relative weights assigned to the explanatory variables, but notice that even in the case $k = 1$, w_t depends nonlinearly on z_t in a manner depending on the scalar value of ξ in that case. Therefore, with a view to optimizing power, we perform the optimization in joint tests even for the case $K = 1$.

The following theorem is used in establishing the limiting distribution of $\hat{S}_B$. Let $C(\Xi)$ denote the metric space of real continuous functions endowed with the uniform metric

$$\sup_{\xi \in \Xi} \| z_1(\xi) - z_2(\xi) \| .$$

Theorem 3.2 *Under H_0 and Assumptions 1-7, $\sqrt{T} s_T(\hat{\theta}, \xi)$, defined in (3.5), converges weakly to a mean-zero Gaussian element $z(\xi)$ of $C(\Xi)$ with covariance function*

$$E\left[z(\xi_1)\, z(\xi_2)' \right] = V\left(\xi_{1,}\xi_2\right)$$

where

$$V(\xi_1, \xi_2) = R(\xi_1, \xi_2) - Q(\xi_1) M^{-1} P(\xi_2)' - P(\xi_1) M^{-1} Q(\xi_2)' + Q(\xi_1) M^{-1} \Sigma M^{-1} Q(\xi_2)' \quad (3.15)$$

and $R(\xi_1, \xi_2) = \lim \frac{1}{T} \sum_{t=1}^{T} E\left[d_t(\theta) d_t(\theta)' w_t(\xi_1)\, w_t(\xi_2) \right]_{\theta = \theta_0},\ R(\xi, \xi) = R(\xi)$.

Note that under the hypothesis of a correctly specified likelihood function, we have the information matrix equality $M = \Sigma$. Therefore, we remark on the possibility that the test might be modified for this restricted version of the null hypothesis by imposing this equality in the variance formula. However, this is not an option we shall consider here.

Since $\sup_{\xi \in \Xi}(\cdot)$ is a continuous functional of $\sqrt{T} s_T(\hat{\theta}, \xi)$, it follows by the continuous mapping theorem that under H_0

$$\hat{S}_B \xrightarrow{d} \sup_{\xi \in \Xi} z(\xi)' V(\xi)^{-1} z(\xi) .$$

The limiting distribution of the joint test statistic $\hat{S}_B$ depends on the data generation process and the specification under the null and thus critical values have to be tabulated for each DGP and estimation model which is unfeasible given the general framework of our test statistic. However, an approximate limiting distribution can be obtained by applying the following approach of Bierens (1990).

Lemma 3.3 *Under Assumptions 1-7, choose independently of the data $\gamma > 0$, $0 < \rho < 1$ and $\xi_0 \in \Xi$. Let $\hat{\xi} = \arg\max_{\xi \in \Xi} S_B(\xi)$ and*

$$\tilde{\xi} = \begin{cases} \xi_0 \text{ if } \hat{S}_B - S_B(\xi_0) \leq \gamma T^{\rho} \\ \hat{\xi} \text{ if } \hat{S}_B - S_B(\xi_0) > \gamma T^{\rho} \end{cases} \quad (3.16)$$

Then, under H_0, $\tilde{S}_B = S_B(\tilde{\xi})$ will have an asymptotic χ^2 distribution with p degrees of freedom, whereas under H_1, $\tilde{S}_B / T \to \sup_{\xi \in \Xi} q(\xi)$ a.s. as $T \to \infty$, where $\sup_{\xi \in \Xi} q(\xi) > 0$.

The approach of basing the test on the pair of statistics $S_B(\xi_0)$ and $\sup_{\xi} S_B(\xi)$, depending on the discrimination device in (3.16) offers the real attraction of being able to use a standard table for implement the test. Alternative methods such as the bootstrap, although feasible, are clearly unattractive for routine applications. An alternative to the formulation in (3.14) is the integrated moment test investigated by Bierens and Ploberger (1997). This involves constructing the statistic as the integral of the function $S_B(\xi)$ with respect to a suitable measure defined for ξ. This approach clearly deserves consideration, but the computational overhead of implementing such procedures by the bootstrap appear to us to make it unlikely such methods will find favour with practitioners, however cheap computing power may become.

In addition to this test of joint restrictions, there are also various ways of examining the information contained in the indicator to yield consistent tests. In general, a principle we could adopt is to construct a one degree of freedom test based on a linear combination of individual components of the indicator vector $s_T(\hat{\theta}, \xi)$ in (3.5). This approach may prove to give power in particular directions. For fixed $\xi \in \Xi$, and a vector of weights $\eta \in \mathbb{R}^p$, a composite test statistic can be constructed as

$$S_{Bc}(\xi, \eta) = \frac{\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \eta' d_t(\hat{\theta}) w_t(\xi) \right)^2}{\eta' \hat{V}(\xi) \eta}, \tag{3.17}$$

where $\eta \in \mathcal{H} = \{\eta \in \mathbb{R}^p : \|\eta\| = 1\}$ without loss of generality, since any scale factor cancels in the ratio. If the interest is in distinguishing between different types of misspecification (e.g., misspecification that occurs in the mean or variance equations of a regression model), individual test statistics can be constructed as a special case of (3.17) by setting η to a column of I_p. In this case, the test statistic is constructed as

$$S_{Bi}(\xi) = \frac{\frac{1}{T} \left(\sum_{t=1}^{T} d_{t,i}(\hat{\theta}) w_t(\xi) \right)^2}{\hat{V}_{ii}(\xi)} \tag{3.18}$$

where $d_{t,i}(\hat{\theta}) = \left. \frac{\partial \ln f(y|x, \theta)}{\partial \theta_i} \right|_{\theta = \hat{\theta}}$, for $i = 1, ..., p$, and $\hat{V}_{ii}(\xi)$ is the ith diagonal element of $\hat{V}(\xi)$ given in (3.12). Thus, the individual tests are defined for $i = 1, \ldots, p$ as

$$\hat{S}_{Bi} = \sup_{\xi \in \Xi} S_{Bi}(\xi). \tag{3.19}$$

The limiting distributions of tests specified by (3.17) is given in the following Theorem.

Theorem 3.3 *Under Assumptions 1-7, for every $\xi \in \mathbb{R}^K / B_0 \cup B^*$, where B_0 is the set defined in Lemma (3.1) for the case $\theta = \theta_0$ and B^* is the set defined in Assumption 7, and $\eta \in \mathcal{H}$, the composite test $S_{Bc}(\xi, \eta)$ has a limiting chi-square distribution with one degree of freedom under H_0 in (3.3), whereas under H_1 in (3.4), $S_{Bc}(\xi, \eta)/T \to q(\xi, \eta)$ a.s., where $q(\xi, \eta) > 0$.*

Almost any choice of ξ and η will yield some power to detect misspecification. However, the composite test can be constructed in a similar way to the method proposed in Bierens (1990) leading to the statistic

$$\hat{S}_{Bc} = \sup_{\psi \in \Psi} S_{Bc}(\psi) \tag{3.20}$$

where $\psi = (\xi', \eta')'$, $\Psi = \Xi \times \mathcal{H}$ and $S_{Bc}(\psi) = S_{Bc}(\xi, \eta)$ is defined in (3.17).

The following theorem, analogous to Theorem 3.2, is used to establish the limiting distribution of the test in (3.20). Let $C(\Psi)$ denote the metric space of real continuous functions endowed with the uniform metric $\sup_{\psi \in \Psi} |z_1(\psi) - z_2(\psi)|$.

Theorem 3.4 *Under H_0 and Assumptions 1-7, $\eta' \sqrt{T} s_T(\hat{\theta}, \xi)$, where $s_T(\hat{\theta}, \xi)$ is defined in (3.5) converges weakly to a mean-zero Gaussian element $z(\psi)$ of $C(\Psi)$ with covariance function*

$$E[z(\psi_1) z(\psi_2)] = \eta_1' V(\xi_1, \xi_2) \eta_2$$

where $V(\xi_1, \xi_2)$ is defined in (3.15).

Since $\sup_{\psi\in\Psi}(\cdot)$ is a continuous functional of $\eta'\sqrt{T}s_T(\hat{\theta},\xi)$,

$$\hat{S}_{Bc} \xrightarrow{d} \sup_{\psi\in\Psi} \frac{z(\psi)^2}{\eta' V(\xi)\eta}$$

under H_0, by the continuous mapping theorem. Given that the limiting distributions of the portmanteau test statistic $\hat{S}_{Bc}$ and individual tests $\hat{S}_{Bi}$ are unknown for the general specification framework, an approximate limiting distribution can be obtained by applying the approach of Bierens (1990).

Theorem 3.5 *Under Assumptions 1-7, choose independently of the data* $\gamma > 0$, $0 < \rho < 1$ *and* $\psi_0 \in \Psi$, *where* $\psi = (\xi', \eta')'$. *Let* $\hat{\psi} = \arg\max_{\psi\in\Psi} S_{Bc}(\psi)$ *and*

$$\tilde{\psi} = \begin{cases} \psi_0 \text{ if } \hat{S}_{Bc} - S_{Bc}(\psi_0) \leq \gamma T^{\rho} \\ \hat{\psi} \text{ if } \hat{S}_{Bc} - S_{Bc}(\psi_0) > \gamma T^{\rho} \end{cases} \tag{3.21}$$

Then, under H_0, $\tilde{S}_{Bc} = S_{Bc}(\tilde{\psi})$ *will have an asymptotic* χ^2 *distribution with one degree of freedom, whereas under* H_1, $\tilde{S}_{Bc}/T \to \sup_{\psi\in\Psi} q(\psi)$ *a.s. as* $T \to \infty$, *where* $\sup_{\psi\in\Psi} q(\psi) > 0$.

4 Experimental evidence

4.1 Tests for continuously distributed data in QML estimation

A fundamental application of these tests is to verify the specification of the conditional mean and variance in regression models, and in this context we are not typically interested in a complete specification of the distribution. Hence, it is appropriate to the specification of the test statistic that the variance matrix $V(\xi)$ in (3.8) is constructed without imposing the information matrix equality $M = \Sigma$ in the final term. We do note the possibility of imposing this restriction, although this will create the hazard of a biased test if the restriction is actually incorrect. In this section of the paper we explore only the unrestricted version.

We have carried out Monte Carlo simulations of our procedures in a variety of different models. Our experiments use sample sizes of 100 and 500 observations and each design was carried out with 10,000 replications. For each of our models, the following tests were computed: (i) the regular Bierens residual test in (2.7), computed from appropriately defined residuals; (ii) the joint score test in (3.14), having p degrees of freedom; (iii) the tests on individual score elements defined in (3.19) and (iv) the composite test defined in (3.20). Although these tests can be applied very straightforwardly to system estimation, we confine attention here to the single equation case ($G = 1$), since it is not clear that the considerable computational overhead from larger models will be justified by additional insights. We have studied models of conditionally Gaussian data and discrete data. Estimation is by (quasi-) maximum likelihood in all cases, which in the Gaussian case in particular means that the Gaussian likelihood is optimized with respect to the residual variance as well as the other parameters.

All the generated models incorporate a mean function $m(x_t, \theta)$ that may be either linear or nonlinear in variables and parameters, and some also contain a variance function $h(x_t, \theta)$ to allow for conditional heteroskedasticity. Table 1 shows the specifications of the different cases studied. This mix of models involves either one or two explanatory variables, since we are interested in the effect of optimizing with respect to ξ in both the former and the latter cases. In each model the null hypothesis, the case estimated, is represented by $\delta = 0$. In the experiments x_{1t}, x_{2t}, x_{3t} and ε_t are all generated independently as $N(0, 1)$. Models M5 and M6 incorporate a threshold effect under the alternative, with parameter values dependent on the sign of a third explanatory

	$m(x_t, \theta)$	$h(x_t, \theta)$
M1:	$\beta_0 + \beta_1 x_{1t} + \delta x_{1t}^2$	σ^2
M2:	$\beta_0 + \beta_1 x_{1t}$	$\exp(\delta x_{1t})^{1/2}$
M3:	$\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \delta x_{1t} x_{2t}$	σ^2
M4:	$\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t}$	$\exp(\delta\,(\alpha_1 x_{1t} + \alpha_2 x_{2t}))^{1/2}$
M5:	$\beta_0 + \beta_1 x_{1t} + \delta\,(\alpha_0 + \alpha_1 x_{1t})\,1\,(x_{3t} < 0)$	σ^2
M6:	$\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} +$ $\delta\,(\alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t})\,1\,(x_{3t} < 0)$	σ^2
M7:	$\beta_0 + \beta_1 x_{1t} + \delta x_{1t}^2$	$\left(\alpha_0 + \alpha_1 x_{1t}^2\right)^{1/2}$
M8:	$\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \delta x_{1t} x_{2t}$	$\left(\alpha_0 + \alpha_1 x_{1t}^2 + \alpha_2 x_{2t}^2\right)^{1/2}$

Table 1: Models of Mean and Variance

variable. Models M7 and M8 are distinctive in featuring conditional heteroskedasticity even under the null hypothesis. Note that these are all random regressor models, such that new samples are generated randomly for each Monte Carlo replication.

The sup-tests involve optimizing the statistic over the hypercubes Ξ or Ψ, of dimension $D = k$ and $D = k + p$ respectively. We employed a simple random search algorithm that does not require differentiability or any smoothness properties of the criterion function. Given a factor a, a collection of $N = aD$ uniformly distributed parameter points are drawn from the current search region, initially chosen as Ξ or Ψ. The function values are ranked, the smallest $N/2$ values discarded and the search region is then shrunk to the smallest hypercube containing the remaining points. The factor a is chosen flexibly, depending on the diameter of the current search region, within the bounds $2.5 < a \leq 10$. The step is repeated until the diameter of the search region does not exceed 10^{-4} to provide a workable trade-off between evaluation speed and required accuracy.

A critical choice in the construction of these tests is the values of the sensitivity parameters γ and ρ defined in Lemma 3.3. As a preliminary, we conducted a detailed comparison of alternative choices using one of our models as the test case. This is Model 3 as defined in Table 1. These experiments were conducted using the identical random numbers to generate the data, to ensure a precise comparison between cases. A selection of these results (corresponding to the best γ found for each of four values of ρ) are presented in Table 3. The choice is not clear-cut, and ideally we should experiment with a larger range of models and sample sizes to form a clear idea of the trade-offs involved. However, on the basis of the comparisons we have tentatively used values of $\gamma = 2$ and $\rho = 0.5$ throughout the main body of experiments that follow.

The results of our experiments with these models are shown in Tables 4 to 8. The column headings are as follows. $\hat{B}$ denotes the appropriate variant of Bierens' original test, in other words, the M-test performed on the covariance of the model residuals and the weight function. $\hat{S}_B$ denotes the joint test on the scores, having p degrees of freedom, whereas $\hat{S}_{Bc}$ is the composite test (sup-test) having 1 degree of freedom. The other columns relate to the 1-degree of freedom tests based on the individual elements of the score. Rejection frequencies when the null hypothesis is true are shown in boldface, and in these cases the test critical values are taken from the relevant chi-squared table. Rejection frequencies when the null hypothesis is false, in normal face, are calculated using critical values from the empirical distributions obtained from the simulations of the null, and hence these are estimates of the true powers. Refer in most cases to Table 1 to find the mean and variance models represented in each row of the tables.

Consider the linear/non-linear regression model with possible heteroskedasticity,

$$y_t = m(x_t, \theta) + h\left(x_t, \theta\right)^{1/2} \varepsilon_t, \tag{4.1}$$

Note that in this general set-up, elements of the vector x_t might explain either or both of mean and variance, and elements of θ could appear in either or both functions likewise. The conditional Gaussian quasi log-likelihood function for the model in (4.1) is

$$L_T\left(\theta\right) = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left(h_t\right) + \frac{\varepsilon_t^2}{h_t}\right] \tag{4.2}$$

with the typical term in the score vector given by

$$d_t\left(\theta\right) = -\frac{1}{2}\left(2\frac{\varepsilon_t}{h_t}\frac{\partial\varepsilon_t}{\partial\theta} - \left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{1}{h_t}\frac{\partial h_t}{\partial\theta}\right). \tag{4.3}$$

where $h_t = h\left(x_t, \theta\right)$. There are many data generation processes, not necessarily Gaussian, for which the criterion function in (4.2) yields consistent and asymptotically normal estimates. This is therefore a case where we need to distinguish between strictly correct specification and our characterization of the null hypothesis. All that matters is the existence of θ_0 satisfying the conditions of the null and containing economically interpretable parameters.

The results for these models, with β_0, β_1, β_2 and σ^2 all equal to 1 are presented in Table 4. We also set α_0, α_1, α_1 and α_2 all equal to 1 in models M4, M7 and M8. Models M5 and M6 incorporate a threshold effect under the alternative and we set $\alpha_0 = -2$, $\alpha_1 = -3$ and $\alpha_2 = -1$. Note that model M1 represents the null hypothesis for models M2 and M5, with $\delta = 0$, and similarly with respect to models M3, M4 and M6. While M1-M6 are linear regressions when $\delta = 0$, and could have been estimated by least squares, we have nonetheless performed all estimations by Gaussian ML so that the variance parameter is estimated and contributes to a score element. Observe that in a regression model, the regular Bierens (1990) test corresponds asymptotically to the score-based test for the intercept. In this context, it is therefore merely one of the several options that we compare for power. Models M7 and M8 are nonlinear under the null hypothesis and contain extra parameters; in these cases the variance intercept α_0 is entered in the column headed $\hat{S}_{B,\sigma^2}$ in the obvious way.

Under the null hypothesis of a linear model, the empirical size of the Bierens test is close to the nominal size of 5%, with the exception of the case when the errors are heteroskedastic and $T = 100$ in which case Bierens test is undersized. The joint test $\hat{S}_B$ is worst-sized of our score tests but the composite test is an improvement in this regard. The individual tests corresponding to the variance component are slightly oversized for $T = 100$ observations, but they are correctly sized (to within experimental error) when $T = 500$, although slight over-rejections still occur when the errors are heteroskedastic. Under the alternative nonlinear model with homoskedastic error terms (models M1 and M3), all test statistics have good comparable size-adjusted power even for $T = 100$. When heteroskedasticity is neglected but the conditional mean is correctly specified, the Bierens test has no power since it it essentially a test of functional form of the mean equation. However, the score-based tests are able to detect this misspecification with the composite test attaining a simulated size-adjusted power of 94.24% for $T = 100$. Moreover, the tests on individual parameters are able to disentangle different sources of misspecification. For example, the statistic corresponding to the variance in regression models is an excellent indicator of heteroskedasticity. Our score-based tests also have very good power in detecting threshold effects, while Bierens test appears insensitive to this misspecification in the mean. When the errors are heteroskedastic under the null hypothesis, such as in the models M7 and M8, Bierens test is not able to detect neglected non-linearity in the mean equation when the number of regressors is two even for a sample size of 500 observations, whereas the score-based tests we propose have good empirical power in these cases.

4.2 Tests in GMM estimation

In this section we consider a model defined by a scalar function $g_t(\theta) = g(y_t, x_t, \theta)$ where the true value of the parameters are defined as solutions to

$$E(g_t(\theta_0)|z_t) = 0 \text{ a.s.} \tag{4.4}$$

In particular, y_t may denote a G-vector of non-exogenous variables, with $G > 1$. In this framework we shall estimate θ_0 by the GMM estimator

$$\hat{\theta} = \arg\min_{\theta\in\Theta} g(\theta)' Z(Z'WZ)^{-1}Z'g(\theta)$$

where $g(\theta) = (g_1(\theta), \ldots, g_T(\theta))'$ and $Z = (z_1, \ldots z_T)'$, and W is a $T \times T$ weighting matrix which for optimally should be set to $E\left[g(\theta_0)\, g(\theta_0)'\right]$. The analogues of the score contributions in the QML estimation are the array elements

$$d_{Tt}(\theta) = D(\theta)' Z\left(Z'WZ\right)^{-1} z_t g_t(\theta)$$

where

$$D(\theta) = \frac{\partial g(\theta)}{\partial \theta'}\ (T \times p),$$

and $\sum_{t=1}^{T} d_{Tt}(\hat{\theta}) = 0$ by construction. We construct our tests by weighting the elements of the sum just as in the preceding section. Whereas the null hypothesis is represented by (4.4) so that it might be natural to base the test on the elements $z_t g_t(\theta)$, when the model is overidentified, it is not the case that $\sum_{t=1}^{T} z_t g_t(\hat{\theta}) = 0$. On the other hand, we can in general only assert that

$$E(d_{Tt}(\theta_0)|z_t) \to 0 \text{ a.s.}$$

as $T \to \infty$. We proceed on the assumption that the resulting size distortions in small samples are of small order, and hence acceptable, under the usual regularity conditions. An advantage of maintaining a common framework with the QML based tests is that the asymptotic derivations in Section 3 go through unamended. It also an advantage to be able to associate a statistic with each parameter in the model, as before. Note that in the just-identified case the tests based on $d_{Tt}(\hat{\theta})w_t$ and $z_t g_t(\hat{\theta})w_t$ are asymptotically equivalent.

A customary test of specification in GMM estimation is the so-called Sargan-Hansen test of overidentification (see Sargan 1964, Hansen 1982) based on the distribution of $z_t g_t(\hat{\theta})$. As in the experiments of the preceding section we used samples of 100 and 500 observations and each design is carried out with 10000 replications. We compare the performance of these tests for the simultaneous equations models specified in Table 2 with M13 being the null model and where we set the values of the parameters in the equation for y_{1t} and β_0 and β_1 in the equations for y_{2t} all equal to 1. The results of these experiments are reported in Table 5, where in all cases we set $W = I_T$. Under the null hypothesis corresponding to model M13, the tests have empirical size close to the nominal size, although the Sargan-Hansen test is slightly over-sized even for 500 observations. In terms of the power properties, while Sargan-Hansen tests might be expected to detect neglected non-linearity and misspecification of the functional form, in practice they have no power in those directions, whereas the Bierens type test and score-based tests have good empirical power with the latter having higher rejection frequencies than the former.

4.3 Probit and logit models

Discrete choice models are typically different from those considered above in the sense that model specification is all-or-nothing business. Either all aspects of the distribution are correctly

Equation 1	
M13-M16:	$y_{1t} = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \varepsilon_{1t}$
Equation 2	
M13:	$y_{2t} = \beta_0 + \beta_1 y_{1t} + \frac{1}{2}(\varepsilon_{1t} + \varepsilon_{2t})$
M14:	$y_{2t} = \beta_0 + \beta_1 y_{1t}^2 + \frac{1}{2}(\varepsilon_{1t} + \varepsilon_{2t})$
M15	$y_{2t} = \beta_0 + \beta_1 y_{1t} + \delta x_{1t}^2 + \frac{1}{2}(\varepsilon_{1t} + \varepsilon_{2t})$
M16:	$\ln y_{2t} = \frac{1}{4}(\beta_0 + \beta_1 y_{1t} + \frac{1}{2}(\varepsilon_{1t} + \varepsilon_{2t}))$

Table 2: Simultaneous Equations Models

specified, or, in general estimator consistency fails. There is no 'quasi-maximum likelihood' for these cases. Although our tests have the same structure as before, there is a crucial difference in the interpretation. The conditional mean of the scores is directly connected with the form of the distribution and hence the latter is amenable to test.

In a generalization of the standard probit and logit models to allow for heteroskedasticity in the latent model, we consider an underlying latent equation with the form

$$y_t^* = m(x_t, \theta) + h(x_t, \theta)^{1/2} \varepsilon_t \tag{4.5}$$

where ε_t is an independent and identically distributed shock with distribution function $F(z) = P(\varepsilon_t < z)$. In the Probit case $F(z) = \Phi(z)$, the standard Gaussian c.d.f., while in the logit model $F(z) = 1/(1 + e^z)$. The binary observed variable is then defined as

$$y_t = \begin{cases} 0 \text{ if } y_t^* \leq 0 \\ 1 \text{ if } y_t^* > 0 \end{cases} \tag{4.6}$$

Let

$$m^*(x_t, \theta) = \frac{m(x_t, \theta)}{h(x_t, \theta)^{1/2}}, \tag{4.7}$$

while noting that when $h_t = \sigma^2$, not depending on t, m_t^* becomes m_t with σ^2 unidentified and hence set to 1. This is the case with all our null hypotheses, for which the statistic is calculated, although we retain the general notation below for consistency. The probabilities are accordingly defined as

$$\begin{aligned} \Pr(y_t = 1 | x_t) &= \Pr(\varepsilon_t > -m^*(x_t, \theta)) \\ &= F(m^*(x_t, \theta)). \end{aligned}$$

The conditional log-likelihood function is

$$\log L_T(\theta) = \sum_{t=1}^{T} y_t \log[F(m^*(x_t, \theta))] + (1 - y_t) \log[1 - F(m^*(x_t, \theta))]$$

and the score contribution takes the form

$$d_t(\theta) = (y_t - F(m^*(x_t, \theta))) q(x_t, \theta) \frac{\partial m^*(x_t, \theta)}{\partial \theta} \tag{4.8}$$

where

$$q(x_t, \theta) = \frac{f(m^*(x_t, \theta))}{F(m^*(x_t, \theta))[1 - F(m^*(x_t, \theta))]}$$

and $f(z) = \partial F(z)/\partial z$. In the logit case, $q(x_t, \theta) = 1$.

Correct specification of the probit/logit models requires that under the null hypothesis

$$\Pr(y_t = 1|x_t) = F(m^*(x_t, \theta_0))$$

and hence (recalling that elements of z_t not included in x_t are irrelevant by hypothesis)

$$\Pr(E(d_t(\theta_0)|z_t) = 0) = 1.$$

The Bierens test (1990) is not designed for discrete choice models but a consistent test statistic can in fact be constructed by generalizing his approach. This involves replacing the regression residuals in the test indicator (2.3) by the generalized residuals, defined for binary choice models as

$$\hat{\varepsilon}_t^* = \left(y_t - F(m^*(x_t, \hat{\theta}))\right) q\left(x_t, \hat{\theta}\right). \tag{4.9}$$

This test therefore differs from the test based on (3.5) by the replacement of the factors $\partial m^*(x_t, \theta)/\partial\theta$ by unity in the terms in the sum.

The results for the probit and logit models are reported in Tables 6 and 7 respectively. Refer to Table 1 for the mean and variance functions with β_0 equal to 0, and $\beta_1 = \beta_2 = 1$, and note that $m = m^*$ for each of the null hypotheses tested, although the data were of course generated using nonlinear latent models M2 and M4 according to (4.5). Models M9 and M10 in Table 6 are new cases, defined by use of a non-Gaussian distribution featuring skewness to generate the binary responses. We used a centred chi-squared with four degrees of freedom to generate the series in these cases, with the mean functions given by models M1 and M3, respectively with $\delta = 0$. The Bierens type test based on the generalized residuals and our score-based tests have good size properties both for the probit and logit models, with the exception of the joint test which is again slightly oversized for both 100 and 500 observations. The tests perform well in detecting neglected nonlinearity, heteroskedasticity and misspecification of the distribution function, with the composite test having overall the best empirical power among the other statistics.

4.4 Count data models

Consider the Poisson model of count data y_t, taking nonnegative integer values, where

$$P(y_t|x_t) = \frac{\exp(-\phi_t)\phi_t^{y_t}}{y_t!}, \text{ for } y_t = 0, 1, 2, ..$$

The Poisson regression specification considered is

$$\ln \phi_t = m^*(x_t, \theta)$$

where m^* is defined by (4.7). With this specification

$$E[y_t|x_t] = Var[y_t|x_t] = \phi_t$$

and the unknown parameter vector θ_0 can be estimated by MLE. Given an independent sample of size T, the log-likelihood function is

$$\ln L = \sum_{t=1}^{T} [-\phi_t + y_t m^*(x_t, \theta) - \ln(y_t!)]$$

and the score contribution is

$$d_t(\theta) = (y_t - \phi_t)\frac{\partial m^*(x_t, \theta)}{\partial\theta}.$$

Therefore, a consistent specification test can be constructed based on the test indicator.

$$s_T(\hat{\theta}, \xi) = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\phi}_t) w\left(x_t, \xi\right) \frac{\partial m^*(x_t, \hat{\theta})}{\partial \theta}.$$

As pointed out by Hausman, Hall and Griliches (1984) and Cameron and Trivedi (1986, 1998), the fact that the mean of the Poisson dependent variable equals its variance is a potentially unrealistic feature of the observed data. Cameron and Trivedi (1986) consider two forms of negative binomial model that arise from a natural generalization of cross-section heterogeneity. In the so-called Negative Binomial 1, the variance of y_t has the specification $var[y_t|x_t] = \phi_t (1 + \alpha)$ and in the Negative Binomial 2 form, $var[y_t|x_t] = \phi_t + \alpha\phi_t^2$.

The log-likelihood function, with the parameterization $\alpha_t = \phi_t/\alpha$ for the Negative Binomial 1 and $\alpha_t = 1/\alpha$ for the Negative Binomial 2, is

$$L_T(\varphi) = \sum_{t=1}^{T} \left[\ln\Gamma(y_t + \alpha_t) - \ln\Gamma(1 + y_t) - \ln\Gamma(\alpha_t) + \alpha_t \ln\left(\frac{\alpha_t}{\alpha_t + \phi_t}\right) + y_t \ln\left(\frac{\phi_t}{\alpha_t + \phi_t}\right) \right]$$

where $\varphi = (\theta', \alpha)'$ and the score vector is

$$\begin{aligned} d_t(\varphi) &= (y_t - \phi_t) \left(\frac{\alpha_t}{\alpha_t + \phi_t}\right) \frac{\partial m(x_t, \theta)}{\partial \theta} \\ &\quad + \left[(\ln\Gamma)'(y_t + \alpha_t) - (\ln\Gamma)'(\alpha_t) + \ln\left(\frac{\alpha_t}{\alpha_t + \phi_t}\right) - \frac{y_t - \phi_t}{\alpha_t + \phi_t} \right] \frac{\partial \alpha_t}{\partial \varphi} \end{aligned}$$

where $(\ln\Gamma)'$ represents the first derivative of the log of the gamma function.

The results of tests on models M1-M4 with parameter values as in Section 4.1 are shown in Table 8 for the Poisson model, and in Table 9 for the Negative Binomial 1 case, where the additional parameter α is equal to 2. Our tables also report the Bierens test where the residual in this case is computed as $\hat{\varepsilon}_t = y_t - \hat{\phi}_t$. The results in Table 8 for the Poisson model suggest that the tests are correctly sized and have good power in detecting nonlinearity and heteroskedasticity. When the null model is the Negative Binomial 1, with the results being reported in Table 9, the tests are slightly oversized for $T = 100$, with the joint test being the worst-sized of all. The tests perform well in detecting nonlinearity and heteroskedasticity for $T = 500$ observations.

As noted in Section 3 above, our test cannot detect the use of an incorrect likelihood function, for example Poisson for Negative Binomial. However, Cameron and Trivedi (1990) suggest a score test for equality of mean and variance, a procedure that could be regarded as complementary to our own.

5 Concluding Remarks

Our reported simulation results show that at least for the given alternatives our tests typically have ample power to detect misspecification. However, the point we wish to emphasize is that these tests are not tailored to the particular model, as is common practice, but apply a single rule to the full range of estimators, and are accordingly very easy to implement routinely.

The other feature that the tables highlight is that the joint chi-square test (having p degrees of freedom) is in general the worst-sized of our alternatives and the so-called composite test (depending on η) improves on the joint test in this regard, as well as having at least equivalent power. The tests on individual parameters are quoted chiefly to see how much information they give on the sources of misspecification. In particular, note that the statistic corresponding to the variance in regression models is an excellent indicator of heteroskedasticity. The so-called regular Bierens test, based on the covariance of residuals with weight functions, should in many cases give a similar result to the individual score based test for the intercept parameter. It is quoted in the tables as a basis for comparison. There are a number of cases where this test has no power in our experiments, for example, regression models with heteroskedasticity in both null and alternative, threshold models and little power in negative binomial models in the context of Poisson model estimation.

In this paper we focus on independently sampled observations. In generalizing our results to time series models, we first note that the likelihood contributions will need to be replaced by conditional contributions where the conditioning variables include lags, similarly to the work of de Jong (1996). However, there is a further condition for correct dynamic specification, that the score contributions, and hence also the terms in our test statistics when suitable defined, should form martingale difference sequences. This could lead to a generalization of the Nyblom-Hansen class of dynamic specification tests (Nyblom 1989, Hansen 1992) for example. However, these important extensions must be left for future research.

Acknowledgement The authors gratefully acknowledge research support from ESRC (Grant Reference RES-000-22-2845).

References

[1] Basmann, R. 1960. On finite sample distributions of generalized classical linear identiability test statistics. Journal of the American Statistical Association 55(292), 650-659.

[2] Bierens, H.J., 1982. Consistent Model Specification Tests. Journal of Econometrics 20, 105-134.

[3] Bierens, H.J., 1990. A consistent conditional moment test of functional form, Econometrica 58, 1443-1458.

[4] Bierens, H.J., and Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests, Econometrica 65, 1129-1151.

[5] Cameron, A.C., and Trivedi, P.K., 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests, Journal of Applied Econometrics 1, 29-53.

[6] Cameron, A.C., and Trivedi, P.K., 1990. Regression-based tests for overdispersion in the Poisson model, Journal of Econometrics 46, 347-64.

[7] Cameron, A.C., and Trivedi, P.K., 1998. Regression Analysis of Count Data, Cambridge, Cambridge University Press.

[8] Chen, X., and Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series, Journal of Econometrics 91, 373-401.

[9] Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians, Oxford, Oxford University Press.

[10] Davidson, J., 2000. Econometric Theory, Oxford, Blackwell Publishers.

[11] Davidson, R., and J.G. MacKinnon, 1981. Several tests of model specification in the presence of alternative hypotheses, Econometrica 49, 781-93.

[12] Davidson, R., and J.G. MacKinnon, 1984. Convenient specification tests for logit and probit models, Journal of Econometrics 25, 241-62.

[13] Delgado, M.A., Dominguez, M.A., and Lavergne, P., 2006. Consistent tests of conditional moment restrictions. Annales d'Economie et de Statistique 81, 33-67.

[14] Durbin, J., 1954. Errors in variables, Review of the International Statistical Institute 22, 23-32.

[15] de Jong, R.M., 1996. The Bierens test under data dependence. Journal of Econometrics 72, 1-32.

[16] Engle, R.F., Henry, D.F., and Richard, J.F., 1983. Exogeneity, Econometrica 51, 277-302.

[17] Escanciano, J.C., 2007. Model checks using residual marked empirical processes. Statistica Sinica 17, 115-38.

[18] Eubank, R.L., and Spielgelman, C.H., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques, Journal of the American Statistical Association 85, 387-92.

[19] Fan, Y., and Li, Q., 1996a. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865-90.

[20] Fan, Y., and Li, Q., 1996b. Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. Unpublished manuscript. Department of Economics, University of Windsor.

[21] Hansen, B. E. 1992. Testing for parameter instability in linear models. Journal of Policy Modelling 14, 517-533.

[22] Hansen, L., 1982. Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

[23] Härdle, W., and Mammen, E., 1993. Comparing nonparametric versus parametric regression fits, Annals of Statistics 21, 1926-47.

[24] Hausman, J.A., 1978. Specification tests in Econometrics, Econometrica 46, 1251-72.

[25] Hausman, J.A., Hall, B., and Griliches, Z., 1984. Economic models for count data with an application to patents-R&D relationship, Econometrica 52, 1984, 909-938.

[26] Hong, Y., and White, H., 1996. Consistent specification testing via nonparametric series regressions. Econometrica 63, 1133-60.

[27] Newey, W.K., 1985. Maximum likelihood specification testing and conditional moment test. Econometrica 53, 1047-1070.

[28] Newey, W.K., 1991. Uniform convergence in probability and stochastic equicontinuity, Econometrica 59, 1161-1167

[29] Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223-230

[30] Ramsey, J.B., 1969. Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society, Series B 31, 350-371.

[31] Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables, Econometrica 26, 393-415.

[32] Sargan, J.D., 1964. Wages and prices in the United Kingdom: a study in econometric methodology, in P. E. Hart, G. Mills and J.K. Whitaker (eds.), Econometric Analysis for National Economic Planning, Butterworth, 25-54

[33] Stinchcombe, M.B., and White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative, Econometric Theory 14, 295-325.

[34] Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models, Journal of Econometrics 30, 415-443.

[35] Whang, Y-J., 2000. Consistent bootstrap tests of parametric regression functions, Journal of Econometrics 98, 27-46.

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

[37] White, H., 1980. A heteroskedastcity-consistent covariance matrix estimator and a direct test of heteroskedasticity, Econometrica 48, 817-38.

[38] White, H., 1982. Maximum likelihood estimation of misspecified models, Econometrica 50, 1-26.

[39] Wu, D-M., 1973. Alternative tests of independence between stochastic regressors and disturbances, Econometrica 41, 733-50.

[40] Zheng, J.X., 1996. A consistent tets of functional form via nonparametric estimation techniques, Journal of Econometrics 75, 263-89.

A Appendix

Proof of Lemma 3.1. The proof follows trivially from Lemma 1 of Bierens (1990). ■

Lemma A.1 *Under Assumptions 1-4*

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}d_t(\theta)d_t(\theta)' - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}d_t(\theta)d_t(\theta)'\right]\right\| = o_p(1) \tag{A-1}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}w_t d_t(\theta) - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}w_t d_t(\theta)\right]\right\| = o_p(1) \tag{A-2}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}w_t d_t(\theta)d_t(\theta)' - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}w_t d_t(\theta)d_t(\theta)'\right]\right\| = o_p(1) \tag{A-3}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}w_t^2 d_t(\theta)d_t(\theta)' - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}w_t^2 d_t(\theta)d_t(\theta)'\right]\right\| = o_p(1) \tag{A-4}$$

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial d_t(\theta)}{\partial\theta'} - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial d_t(\theta)}{\partial\theta'}\right]\right\| = o_p(1) \tag{A-5}$$

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(w_t\frac{\partial d_t(\theta)}{\partial\theta'}\right) - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}\left(w_t\frac{\partial d_t(\theta)}{\partial\theta'}\right)\right]\right\| = o_p(1) \tag{A-6}$$

Proof of Lemma A.1. Under Assumptions 1-4, the uniform convergence results follow by applying a uniform law of large numbers (ULLN) for independent, not identically distributed (i.n.i.d.) random variables (e.g. White (1980), Lemma 2.3). For a generic function $q_t(\theta,\xi)$ in order to show that

$$\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}q_t(\theta\,\xi) - \lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}q_t(\theta,\xi)\right]\right\| = o_p(1),$$

it is sufficient to establish that $E\sup_{\theta\in\Theta,\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}q_t(\theta\,\xi)\right\|^{1+s} < \infty$ uniformly in t for some $s > 0$. For example, (A-1) follows by the Cauchy-Schwartz inequality and Assumption 4(i). The other parts of the Lemma follow similarly from Assumption 4(i)-(iv). ■

Proof of Lemma 3.2. A mean value expansion of $\sqrt{T}s_T(\hat\theta,\xi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_t(\hat\theta)w_t$ about the true parameter θ_0 yields

$$\sqrt{T}s_T(\hat\theta,\xi) = \sqrt{T}s_T(\theta_0,\xi) - \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial d_t\left(\bar\theta_{i,\xi}\right)}{\partial\theta'}w_t\right)\sqrt{T}(\hat\theta-\theta_0)$$

where $\bar\theta_{i,\xi}$ is a mean value, in general different for each component of the score vector, such that $\left\|\bar\theta_{i,\xi}-\theta_0\right\| \le \left\|\hat\theta-\theta_0\right\| = O_p\left(T^{-1/2}\right)$ by Assumption 6. Under Assumptions 1-6 and employing Lemma A.1, the mean value expansion above becomes

$$\begin{aligned}
\sqrt{T}s_T(\hat\theta,\xi) &= \sqrt{T}s_T(\theta_0,\xi) - Q(\xi)\sqrt{T}(\hat\theta-\theta_0) + o_p(1)\\
&= \sqrt{T}s_T(\theta_0,\xi) - Q(\xi)M^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_t(\theta_0) + o_p(1)\\
&= \sqrt{T}z_T(\theta_0,\xi) + o_p(1)
\end{aligned}$$

where M and $Q(\xi)$ are defined in (3.6) and (3.9), respectively, and let

$$z_T(\theta_0,\xi) = \frac{1}{T}\sum_{t=1}^{T} d_t(\theta_0)w_t - Q(\xi)\,M^{-1}\frac{1}{T}\sum_{t=1}^{T} d_t(\theta_0). \tag{A-7}$$

For fixed $\xi \in \mathbb{R}^K$, the Liapounov CLT for i.n.i.d. random variables (see Theorem 23.11, Davidson, 1994) and Assumption 4(i)-(ii) ensure

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d_t(\theta_0)w_t \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d_t(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} R(\xi) & P(\xi) \\ P(\xi)' & \Sigma \end{pmatrix}\right)$$

where $P(\xi), R(\xi)$ and Σ are defined in (3.10), (3.11) and (3.7), respectively.

Therefore,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} d_t(\hat{\theta})w_t \xrightarrow{d} N(0, V(\xi))$$

where

$$V(\xi) = R(\xi) - Q(\xi)\,M^{-1}P(\xi)' - P(\xi)\,M^{-1}Q(\xi)' + Q(\xi)\,M^{-1}\Sigma M^{-1}Q(\xi)'. \tag{A-8}$$

■

Lemma A.2 *Under H_0 and Assumptions 1-6,*

$$\hat{V}(\xi)^{-1/2}\sqrt{T}s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\sqrt{T}z_T(\theta_0,\xi) = o_p(1) \tag{A-9}$$

uniformly over $\xi \in \Xi$, where $z_T(\theta_0,\xi)$ is defined in (A-7).

Proof of Lemma A.2. We have that

$$\begin{aligned}
&\sup_{\xi\in\Xi}\left\|\hat{V}(\xi)^{-1/2}\sqrt{T}s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\sqrt{T}z_T(\theta_0,\xi)\right\| \\
&\le \sup_{\xi\in\Xi}\left\|\hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2}\right\|\sup_{\xi\in\Xi}\left\|\sqrt{T}s_T(\hat{\theta},\xi)\right\| \\
&+\sup_{\xi\in\Xi}\left\|\sqrt{T}s_T(\hat{\theta},\xi) - \sqrt{T}z_T(\theta_0,\xi)\right\|\sup_{\xi\in\Xi}\left\|V(\xi)^{-1/2}\right\|
\end{aligned} \tag{A-10}$$

By Lemmas 3.2 and A.1, and Slutsky's Theorem

$$\sup_{\xi\in\Xi}\left\|\hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2}\right\| = o_p(1).$$

Moreover, by Lemma 3.2

$$\begin{aligned}
\sqrt{T}s_T(\hat{\theta},\xi) &= \sqrt{T}z_T(\theta_0,\xi) + o_p(1) \\
&= O_p(1)
\end{aligned}$$

uniformly over ξ. Therefore,

$$\sup_{\xi\in\Xi}\left\|\hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2}\right\|\sup_{\xi\in\Xi}\left\|\sqrt{T}s_T(\hat{\theta},\xi)\right\| = o_p(1).$$

Now

$$\sup_{\xi\in\Xi}\left\|\sqrt{T}s_T(\hat{\theta},\xi) - \sqrt{T}z_T(\theta_0,\xi)\right\| = o_p(1)$$

by Lemma 3.2 and since $\sup_{\xi\in\Xi}\left\|V(\xi)^{-1/2}\right\| = O_p(1)$, the second term in the expression (A-10) is $o_p(1)$. This proves the result. ■

Lemma A.3 *Under Assumptions 1-7 and H_1, there exists for each $\xi \in \mathbb{R}^K$ some function $\pi_\xi : \mathbb{R}^p \to \mathbb{R}^p$ such that*

$$\hat{V}(\xi)^{-1/2} s_T(\hat{\theta}, \xi) - V(\xi)^{-1/2}\pi_\xi = o_p(1)$$

where $V(\xi)^{-1/2}\pi_\xi \neq 0$ for all $\xi \in \mathbb{R}^K$ except possibly in a set of Lebesgue measure zero.

Proof of Lemma A.3. We can write for each $\xi \in \mathbb{R}^K$

$$\begin{aligned}\left\|\hat{V}(\xi)^{-1/2} s_T(\hat{\theta}, \xi) - V(\xi)^{-1/2}\pi_\xi\right\| &\leq \left\|\hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2}\right\| \|\pi_\xi\| \\ &+ \left\|s_T(\hat{\theta}, \xi) - \pi_\xi\right\| \left\|\hat{V}(\xi)^{-1/2}\right\|. \end{aligned} \tag{A-11}$$

For the second right-hand side term, $\hat{V}(\xi)^{-1/2} = O_p(1)$ and Lemma A.1(A-2) establishes that

$$\operatorname*{plim}_{T\to\infty} \sup_{\theta\in\Theta} \left\| s_T(\theta, \xi) - \lim_{T\to\infty} E\left[s_T(\theta, \xi)\right]\right\| = 0.$$

Therefore, set $\pi_\xi = \lim_{T\to\infty} E\left[s_T(\theta_1, \xi)\right]$, where $\theta_1 = \operatorname{plim}\hat{\theta}$ under H_1. Moreover, in the first term $\hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2} = o_p(1)$ by Lemma A.1 and Slutsky's Theorem and since $\|\pi_\xi\| = O(1)$ by Assumption 4(ii), the first term on the right-hand side of (A-11) is $o_p(1)$. Therefore, it has been established that

$$\left\|\hat{V}(\xi)^{-1/2} s_T(\hat{\theta}, \xi) - V(\xi)^{-1/2}\pi_\xi\right\| = o_p(1)$$

Now by Assumption 7 and Lemma 3.1, $V(\xi)^{-1/2}\pi_\xi \neq 0$ for every $\xi \in \mathbb{R}^K/B$. ■

Proof of Theorem 3.1. The result under H_0 follows from Lemmas 3.2 and A.2. Under H_1, it follows from Lemma A.3 that $\operatorname{plim}_{T\to\infty} S_B/T = \pi'_{\xi 0} V(\xi)^{-1}\pi_{\xi 0} = \rho(\xi)$, where $\pi_{\xi,0} = \lim_{T\to\infty} E\left[s_T(\theta_0, \xi)\right] = 0$ only on a set B_0 of Lebesgue measure zero defined in Lemma 3.1. Therefore, $P\left[\rho(\xi) > 0\right] = 1$ for each $\xi \in \mathbb{R}^K/B_0$.

Lemma A.4 *Under Assumptions 1-4 and H_0, $\sqrt{T} z_T(\theta_0, \xi)$ defined in (A-7) is tight in Ξ.*

Proof of Lemma A.4. Consider $\lambda \in R^p$ such that $\lambda'\lambda = 1$. Following Newey (1991, p1163), in order to show that $\sqrt{T} z_T(\theta_0, \xi)$ is tight in Ξ, it suffices to prove that
(i) For each $\delta > 0$ and $\xi_0 \in \Xi$ there exists an ε such that

$$P\left[\left|\sqrt{T}\lambda' z_T(\theta_0, \xi_0)\right| > \varepsilon\right] \leq \delta$$

for all $t \geq 1$.
(ii) For each $\delta > 0$ and $\varepsilon > 0$ there exists $\alpha > 0$ such that

$$P\left[\sup_{\|\xi_1 - \xi_2\| < \alpha} \left|\lambda'\left(\sqrt{T} z_T(\theta_0, \xi_1) - \sqrt{T} z_T(\theta_0, \xi_2)\right)\right| \geq \varepsilon\right] \leq \delta$$

for all $T \geq T_0$, where $T < \infty$. The condition (i) follows from Lemma 3.2 which establishes that $\sqrt{T} z_T(\theta_0, \xi_0) = O_p(1)$. To show condition (ii), since $z_T(\theta_0, \xi) = s_T(\theta_0, \xi) - Q(\xi) M^{-1} d_T(\theta_0)$, where $d_T(\theta_0) = T^{-1}\sum_{t=1}^T d_t(\theta_0)$ by the continuity of $Q(\xi)$, then it is sufficient to show that for all $\lambda \in \mathbb{R}^p$ such that $\lambda'\lambda = 1$

$$E\left(\sup_{\|\xi_1 - \xi_2\| < \alpha} \left|\lambda'\left(\sqrt{T} s_T(\theta_0, \xi_1) - \sqrt{T} s_T(\theta_0, \xi_2)\right)\right|\right) < \infty.$$

Notice that

$$
\begin{aligned}
&E\left(\sup_{\|\xi_1-\xi_2\|<\alpha}\left|\lambda'\left(\sqrt{T}s_T(\theta_0,\xi_1)-\sqrt{T}s_T(\theta_0,\xi_2)\right)\right|\right)\\
&\qquad\leq E\left(\sup_{\|\xi_1-\xi_2\|<\alpha}\left|1/\sqrt{T}\sum_{t=1}^{T}\lambda'd_t\left(\theta_0\right)\right|\left|\exp\left(\xi_1'\Phi\left(x_t\right)\right)-\exp\left(\xi_2'\Phi\left(x_t\right)\right)\right|\right).
\end{aligned}
$$

Now

$$
\begin{aligned}
\exp\left(\xi_1'\Phi\left(x_t\right)\right)&=\sum_{i=0}^{\infty}\frac{\left(\xi_1'\Phi\left(x_t\right)\right)^i}{i!}\\
&=\sum_{i=0}^{\infty}\frac{1}{i!}\sum_{m_1,\ldots,m_K=0}^{i}\binom{i}{m_1,\ldots,m_K}\xi_{1,1}^{m_1}\cdots\xi_{1,K}^{m_K}\Phi_1\left(x_t\right)^{m_1}\cdots\Phi_K\left(x_t\right)^{m_K}
\end{aligned}
$$

and thus

$$
\begin{aligned}
&E\left(\sup_{\|\xi_1-\xi_2\|<\alpha}\left|1/\sqrt{T}\sum_{t=1}^{T}\lambda'd_t\left(\theta_0\right)\right|\left|\exp\left(\xi_1'\Phi\left(x_t\right)\right)-\exp\left(\xi_2'\Phi\left(x_t\right)\right)\right|\right)\\
&\quad\leq E\left(\sup_{\|\xi_1-\xi_2\|<\alpha}\left|1/\sqrt{T}\sum_{t=1}^{T}\lambda'd_t\left(\theta_0\right)\right|\right.\\
&\qquad\left.\times\left|\sum_{i=0}^{\infty}\frac{1}{i!}\sum_{m_1,m_2,\ldots,m_K=0}^{i}\binom{i}{m_1,\ldots,m_K}\left(\xi_{1,1}^{m_1}\cdots\xi_{1,K}^{m_K}-\xi_{2,1}^{m_1}\cdots\xi_{2,K}^{m_K}\right)\Phi_1\left(x_t\right)^{m_1}\cdots\Phi_K\left(x_t\right)^{m_K}\right|\right).
\end{aligned}
$$

Now since $\Xi=[-b,b]^K$ and given that $\|\xi_1-\xi_2\|<\alpha$, note that

$$
\begin{aligned}
\left|\xi_{1,1}^{m_1}\cdots\xi_{1,K}^{m_K}-\xi_{2,1}^{m_1}\cdots\xi_{2,K}^{m_K}\right|&=\left|\sum_{j=1}^{K}\left(\xi_{1,j}^{m_j}-\xi_{2,j}^{m_j}\right)\prod_{p=1}^{j-1}\xi_{1,p}^{m_p}\prod_{p=j+1}^{K}\xi_{2,p}^{m_p}\right|\\
&\leq\sum_{j=1}^{K}\left|\xi_{1,j}^{m_j}-\xi_{2,j}^{m_j}\right|\prod_{p=1}^{j-1}\left|\xi_{1,p}^{m_p}\right|\prod_{p=j+1}^{K}\left|\xi_{2,p}^{m_p}\right|\\
&\leq b^{K-1}\alpha^2\\
&<\infty
\end{aligned}
$$

where, in the second and third members, we use the convention that $\prod_{p=a}^{b}\xi_{i,p}^{m_p}=1$ if $a>b$ for $i=1,2$. Finally, since

$$
\sum_{m_1,\ldots,m_K=0}^{i}\binom{i}{m_1,\ldots,m_K}\Phi_1\left(x_t\right)^{m_1}\cdots\Phi_K\left(x_t\right)^{m_K}=\left(\iota'\Phi\left(x_t\right)\right)^i
$$

where $\iota=(1,\ldots,1)'$ is the summation vector, we have

$$
\begin{aligned}
&E\left(\sup_{\|\xi_1-\xi_2\|<\alpha}\left|\lambda'\left(\sqrt{T}s_T(\theta_0,\xi_1)-\sqrt{T}s_T(\theta_0,\xi_2)\right)\right|\right)\\
&\quad\leq b^{K-1}\alpha^2E\left(\left|1/\sqrt{T}\sum_{t=1}^{T}\lambda'd_t\left(\theta_0\right)\right|\left|\sum_{i=0}^{\infty}\frac{\left(\iota'\Phi\left(x_t\right)\right)^i}{i!}\right|\right)\\
&\quad=b^{K-1}\alpha^2E\left(\left|1/\sqrt{T}\sum_{t=1}^{T}\lambda'd_t\left(\theta_0\right)\right|\left|\exp\left(\iota'\Phi\left(x_t\right)\right)\right|\right)
\end{aligned}
$$

$$\leq b^{K-1}\alpha^2\left[E\left(T^{-1/2}\sum_{t=1}^{T}\lambda' d_t(\theta_0)\right)^2\right]^{1/2}E\left[\left(\exp\left(\iota'\Phi(x_t)\right)\right)^2\right]^{1/2}$$
$$<\infty$$

by Assumption 4(i). ■

Proof of Theorem 3.2. The result follows from Lemmas 3.2, A.2 and A.4. ■

Proof of Lemma 3.3. Under H_0, from Theorem 3.1, $\widehat{S}_B - S_B(\xi_0) = O_p(1)$, so for any $\gamma > 0$, $\rho \in (0,1)$, $P\left[\widehat{S}_B(\xi) - S_B(\xi_0) > \gamma T^{\rho}\right] \to 0$ and $\lim_{T\to\infty} P\left[\tilde{\xi} = \xi_0\right] = 1$. Thus, under H_0, the test is asymptotically based on $S_B(\xi_0)$ with probability 1 and, since conditionally on ξ_0, $S_B(\xi_0) \xrightarrow{d} \chi_1^2$, then $\widetilde{S}_B \xrightarrow{d} \chi_1^2$. Under H_1, the asymptotic distribution follows from Theorem 3.1.
■

Lemma A.5 *Under H_1 and Assumptions 1-7, there exists for each $\xi \in \mathbb{R}^K$ and $\eta \in \mathcal{H}$ some function $\pi_\xi : \mathbb{R}^p \to \mathbb{R}^p$ such that*

$$\hat{V}(\xi)^{-1/2}\eta' s_T(\hat{\theta},\xi) - V(\xi)\eta'\pi_\xi = o_p(1)$$

where $V(\xi)\eta'\pi_\xi \neq 0$ for all $\xi \in \mathbb{R}^K$ except possibly in a set of Lebesgue measure zero.

Proof. The proof follows straightforwardly from Lemma A.3 since for each $\xi \in \mathbb{R}^K$, $\eta \in \mathcal{H}$

$$\left\|\hat{V}(\xi)^{-1/2}\eta' s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\eta'\pi_\xi\right\| \leq \left\|\hat{V}(\xi)^{-1/2} - V(\xi)^{-1/2}\right\|\|\eta\|\|\pi_\xi\| + \|\eta\|\left\|s_T(\hat{\theta},\xi) - \pi_\xi\right\|\left\|\hat{V}(\xi)^{-1/2}\right\|$$

where $\|\eta\| = 1$. ■

Proof of Theorem 3.3. The proof under H_0 follows easily from Lemmas 3.2 and A.2, since for each $\xi \in \mathbb{R}^K$ and $\eta \in \mathcal{H}$

$$\begin{aligned}\left|\hat{V}(\xi)^{-1/2}\eta'\sqrt{T}s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\eta'\sqrt{T}z_T(\theta_0,\xi)\right| &\leq \|\eta\|\left\|\hat{V}(\xi)^{-1/2}\sqrt{T}s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\sqrt{T}z_T(\theta_0,\xi)\right\| \\ &= \left\|\hat{V}(\xi)^{-1/2}\sqrt{T}s_T(\hat{\theta},\xi) - V(\xi)^{-1/2}\sqrt{T}z_T(\theta_0,\xi)\right\|\end{aligned}$$

given that $\|\eta\| = 1$.Under H_1, from Lemma A.5, $\operatorname{plim}_{T\to\infty} S_{Bc}/T = \pi_{\xi 0}'\eta\left(\eta' V(\xi)\eta\right)^{-1}\eta'\pi_{\xi 0} = \rho(\xi,\eta)$, where $\eta \neq 0$ since $\|\eta\| = 1$ and $\pi_{\xi,0} = \lim_{T\to\infty} E\left[s_T(\theta_0,\xi)\right] = 0$ only on a set B_0 of Lebesgue measure zero defined in Lemma 3.1. Therefore, $P\left[\rho(\xi,\eta) > 0\right] = 1$ for each $\xi \in \mathbb{R}^K/B_0$ and $\eta \in \mathcal{H}$. ■

Lemma A.6 *Under Assumption 1-4 and H_0, $\eta'\sqrt{T}z_T(\theta_0,\xi)$ is tight in Ψ.*

Proof of Lemma A.6. Similar to Lemma A.4, it suffices to prove that
(i) For each $\delta > 0$ and $\psi_0 \in \Psi$, where $\psi_0 = \left(\xi_0', \eta_0'\right)'$ there exists an ε such that

$$P\left[\left|\eta_0'\sqrt{T}z_T(\theta_0,\xi_0)\right| > \varepsilon\right] \leq \delta$$

for all $t \geq 1$.
(ii) For each $\delta > 0$ and $\varepsilon > 0$ there exists $\alpha > 0$ and $\beta > 0$ such that

$$P\left[\sup_{\|\eta_1-\eta_2\|<\beta,\|\xi_1-\xi_2\|<\alpha}\left|\eta_1'\sqrt{T}z_T(\theta_0,\xi_1) - \eta_2'\sqrt{T}z_T(\theta_0,\xi_2)\right| \geq \varepsilon\right] \leq \delta$$

for all $T \geq T_0$, where $T < \infty$. The condition (i) follows from Lemma 3.2 which establishes that $\sqrt{T} z_T(\theta_0,\xi_0) = O_p(1)$ and thus $\eta_0' \sqrt{T} z_T(\theta_0,\xi_0) = O_p(1)$ since $\|\eta_0\| = 1$. To show condition (ii), notice that

$$\begin{aligned}
\sup_{\|\eta_1-\eta_2\|<\beta, \|\xi_1-\xi_2\|<\alpha} & \left| \eta_1' \sqrt{T} z_T(\theta_0,\xi_1) - \eta_2' \sqrt{T} z_T(\theta_0,\xi_2) \right| \\
& \leq \sup_{\|\eta_1-\eta_2\|<\beta} \|\eta_1 - \eta_2\| \sup_{\xi\in\Xi} \left\| \sqrt{T} z_T(\theta_0,\xi) \right\| \\
& \qquad + \sup_{\|\xi_1-\xi_2\|<\alpha} \left\| \sqrt{T} z_T(\theta_0,\xi_1) - \sqrt{T} z_T(\theta_0,\xi_2) \right\| \sup_{\eta\in\mathcal{H}} \|\eta\| \\
& \leq \beta \sup_{\xi\in\Xi} \left\| \sqrt{T} z_T(\theta_0,\xi) \right\| + \sup_{\|\xi_1-\xi_2\|<\alpha} \left\| \sqrt{T} z_T(\theta_0,\xi_1) - \sqrt{T} z_T(\theta_0,\xi_2) \right\|.
\end{aligned}$$

Now, since $\sup_{\xi\in\Xi} \left\| \sqrt{T} z_T(\theta_0,\xi) \right\| = O_p(1)$ by Theorem 3.2 and $\sup_{\eta\in\mathcal{H}} \|\eta\| = 1$, the result follows by applying condition (ii) of Lemma A.4. ■

Proof of Theorem 3.4. The result follows from Lemmas 3.2 and A.2 and A.6. ■

Proof of Theorem3.5. Under H_0, Theorem 3.3, $\widehat{S}_{Bc} - S_{Bc}(\psi_0) = O_p(1)$, so for any $\gamma > 0$, and $\rho \in (0,1)$, $P\left[\widehat{S}_{Bc}(\psi) - S_{Bc}(\psi_0) > \gamma T^\rho\right] \to 0$ and $\lim_{T\to\infty} P\left[\tilde{\psi} = \psi_0\right] = 1$. Thus, under H_0, the test is asymptotically based on $S_{Bc}(\psi_0)$ with probability 1 and since conditionally on ψ_0, $S_{Bc}(\psi_0) \xrightarrow{d} \chi_1^2$, and ψ_0 is independent of the data generating process, then $\widetilde{S}_{Bc} \xrightarrow{d} \chi_1^2$. Under H_1, the asymptotic distribution follows from Theorem 3.3. ■

γ	ρ	δ	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$	$\hat{S}_{B,\beta_2}$	$\hat{S}_{B,\sigma^2}$
$T = 100$									
8	0.2	0	4.94	9.83	7.62	5.65	6.09	4.47	7.48
			0.02	*0.47*	*2.05*	*.06*	*0*	*0*	*0*
		0.2	23.41	15.96	16.19	24.55	17.29	27.86	4.30
			0.01	*.1.27*	*5.24*	*0.31*	*0.38*	*0.15*	*0.03*
5	0.3	0	4.94	9.87	7.70	5.65	6.09	4.47	7.48
			0.02	*0.51*	*.0214*	*0.06*	*0*	*0.01*	*0.1*
		0.2	23.42	15.92	16.24	24.57	17.29	27.87	4.30
			0.02	*1.37*	*5.50*	*0.33*	*0.38*	*0.16*	*0.03*
2	0.5	0	4.94	9.85	7.68	5.65	6.09	4.47	7.48
			0.02	*0.49*	*2.11*	*0.06*	*0*	*0.01*	*0.1*
		0.2	23.41	15.87	16.15	24.56	17.29	27.87	4.30
			0.01	*10.32*	*50.83*	*0.32*	*0.38*	*0.16*	*0.03*
1	0.7	0	4.92	9.47	6.06	5.60	6.09	4.46	7.40
			0	*0.08*	*0.42*	*0.01*	*0*	*0*	*0*
		0.2	23.40	16.17	17.45	24.60	17.11	17.95	4.40
			0	*.0032*	*.0147*	*.0003*	*.0009*	*.0003*	*.0001*
$T = 500$									
8	0.2	0	5.21	6.60	5.26	5.36	5.27	4.77	5.89
			0	*0.01*	*0.07*	*0*	*0*	*0*	*0*
		0.2	81.75	71.98	76.35	82.02	79.95	80.09	4.70
			0.08	*1.19*	*8.37*	*0.31*	*1.22*	*1.43*	*0*
5	0.3	0	5.21	6.59	5.24	5.36	5.27	4.77	5.89
			0	*0*	*0.05*	*0*	*0*	*0*	*0*
		0.2	81.73	71.75	75.26	81.96	79.81	79.88	4.70
			0.01	*0.41*	*3.59*	*0.08*	*0.26*	*0.38*	*0*
2	0.5	0	5.21	6.59	5.20	5.36	5.27	4.77	5.89
			0	*0*	*0*	*0*	*0*	*0*	*0*
		0.2	81.73	71.62	74.45	81.94	79.79	79.85	4.70
			0	*0.01*	*0.29*	*0*	*0.01*	*0.01*	*0*
1	0.7	0	5.21	6.59	5.20	5.36	5.27	4.77	5.89
			0	*0*	*0*	*0*	*0*	*0*	*0*
		0.2	81.73	71.61	74.35	81.94	79.79	79.85	4.70
			0	*0.01*	*0.29*	*0*	*0.01*	*0*	*0*

Table 3: Rejection frequencies (%) for Model 3 with alternative statistic selection criteria. % of replications in which sup-statistic selected is shown in italics.

Model	δ	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$	$\hat{S}_{B,\beta_2}$	$\hat{S}_{B,\sigma^2}$	$\hat{S}_{B,\alpha_1}$	$\hat{S}_{B,\alpha_2}$
$T = 100$										
M1,2,5	0	**5.40**	**9.14**	**7.03**	**6.23**	**5.70**	**-**	**7.63**	**-**	**-**
M1	0.4	86.51	95.37	89.70	92.44	99.48	-	8.53	-	-
	0.8	99.43	100	99.63	100	99.96	-	16.02	-	-
M2	0.4	4.27	42.02	41.07	4.23	5.36	-	57.85	-	-
	0.8	4.16	96.08	94.24	4.44	8.59	-	98.79	-	-
M5	0.4	4.16	37.68	38.92	4.22	6.54	-	47.01	-	-
	0.8	7.14	99.50	91.07	6.42	6.62	-	99.84	-	-
M3,4,6	0	**4.94**	**9.85**	**7.68**	**5.65**	**6.09**	**4.47**	**7.48**	**-**	**-**
M3	0.4	66.71	57.42	65.34	69.40	72.34	69.96	5.31	-	-
	0.8	95.89	99.22	99.54	99.26	99.60	99.59	13.28	-	-
M4	0.4	5.02	75.16	74.00	6.25	5.81	7.55	89.61	-	-
	0.8	5.54	99.28	99.40	9.79	9.33	13.32	99.95	-	-
M6	0.4	3.53	31.73	32.29	4.56	16.68	7.68	41.23	-	-
	0.8	11.34	83.49	80.87	12.18	6.28	12.38	90.81	-	-
M7	0	**2.50**	**12.32**	**10.34**	**7.40**	**6.75**	**-**	**7.90**	**8.63**	**-**
	0.4	65.70	42.99	44.55	55.00	65.77	-	4.74	4.21	-
	0.8	98.88	94.94	92.03	96.80	99.21	-	3.55	3.61	-
M8	0	**2.22**	**13.14**	**11.52**	**6.43**	**7.54**	**6.05**	**7.06**	**6.84**	**6.30**
	0.4	7.36	10.76	10.48	10.86	11.14	17.40	5.13	4.35	4.47
	0.8	8.17	39.59	41.72	52.50	45.22	46.41	4.51	4.03	4.51
$T = 500$										
M1,2,5	0	**4.76**	**6.14**	**5.60**	**4.92**	**4.94**	**-**	**5.49**	**-**	**-**
M1	0.4	100	100	99.77	100	100	-	36.07	-	-
	0.8	100	100	99.93	100	99.99	-	72.03	-	-
M2	0.4	5.32	99.69	97.73	5.12	5.53	-	99.93	-	-
	0.8	4.71	100	100	4.63	6.71	-	100	-	-
M5	0.4	10.40	99.25	85.56	10.26	6.09	-	99.91	-	-
	0.8	2.42	99.99	99.68	2.51	16.17	-	100	-	-
M3,4,6	0	**5.21**	**6.59**	**5.20**	**5.36**	**5.27**	**4.77**	**5.89**	**-**	**-**
M3	0.4	99.96	99.97	99.94	99.96	99.95	99.99	7.21	-	-
	0.8	100	100	99.97	100	100	100	19.59	-	-
M4	0.4	4.75	99.98	99.91	4.96	5.18	5.65	100	-	-
	0.8	4.58	100	100	5.60	6.76	7.27	100	-	-
M6	0.4	3.16	95.14	61.11	3.20	3.43	9.64	99.04	-	-
	0.8	1.50	100	99.84	1.60	6.28	10.74	100	-	-
M7	0	**4.24**	**7.86**	**6.99**	**5.49**	**5.43**	**-**	**6.01**	**6.06**	**-**
	0.4	99.96	99.83	97.26	99.84	99.97	-	4.47	5.19	-
	0.8	100	99.92	97.10	99.98	99.94	-	6.38	7.96	-
M8	0	**4.53**	**9.46**	**6.27**	**5.67**	**5.38**	**5.49**	**6.07**	**6.87**	**6.93**
	0.4	8.93	43.61	55.94	61.02	62.60	63.44	4.64	4.37	4.83
	0.8	29.29	98.96	99.49	99.49	99.81	99.59	4.15	3.67	3.93

Table 4: Rejection frequencies (%) for Gaussian models (γ=2 and ρ=0.5)

	δ	Sarg	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$
$T = 100$							
M13	0	**7.46**	**6.13**	**6.17**	**4.98**	**6.18**	**4.72**
M14	-	6.00	88.40	91.58	96.90	88.07	97.34
M15	0.4	11.01	5.80	13.37	20.96	5.74	18.81
	0.8	22.15	7.92	24.95	27.85	8.15	26.28
M16	-	2.55	37.14	45.47	62.62	36.53	65.37
$T = 500$							
M13	0	**7.05**	**5.79**	**5.89**	**4.93**	**5.88**	**4.80**
M14	-	2.86	100	100	100	100	100
M15	0.4	6.62	28.19	84.25	86.82	28.02	89.08
	0.8	30.89	47.48	99.53	99.91	47.07	99.67
M16	-	3.51	98.81	99.34	99.39	98.84	99.35

Table 5: Rejection frequencies (%) for GMM models (γ=2 and ρ=0.5)

Model	δ	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$	$\hat{S}_{B,\beta_2}$
$T = 100$							
M1,2,9	0	**7.69**	**7.36**	**6.26**	**7.62**	**5.75**	-
M1	0.4	45.11	38.96	59.02	45.11	60.14	-
	0.8	98.53	96.92	99.07	98.53	99.28	-
M2	0.4	13.78	9.58	13.51	13.78	11.02	-
	0.8	29.92	25.75	36.93	29.93	30.78	-
M9	-	90.52	82.75	80.83	90.53	5.78	-
M3,4,10	0	**6.40**	**6.67**	**5.89**	**6.42**	**4.30**	**8.58**
M3	0.4	13.42	6.64	13.85	13.42	14.65	4.65
	0.8	32.89	19.45	41.18	32.90	38.82	18.85
M4	0.4	13.10	7.59	10.78	13.10	10.55	4.04
	0.8	28.29	22.87	39.38	28.29	34.19	18.09
M10	-	16.13	10.80	20.41	16.12	20.71	7.51
$T = 500$							
M1,2,9	0	**5.43**	**9.07**	**5.51**	**5.43**	**5.74**	**-**
M1	0.4	99.93	99.88	99.69	99.94	99.99	-
	0.8	100	99.99	99.97	99.98	99.98	-
M2	0.4	54.69	38.14	58.07	54.69	49.21	-
	0.8	89.11	92.63	97.25	89.12	95.38	-
M9	-	99.91	99.31	98.39	99.94	96.55	-
M3,4,10	0	**5.09**	**7.40**	**5.82**	**5.09**	**5.72**	**5.77**
M3	0.4	49.60	38.48	59.66	49.60	35.79	38.12
	0.8	98.17	90.18	98.30	98.17	88.22	89.19
M4	0.4	35.07	30.38	58.18	35.07	38.47	37.89
	0.8	74.81	88.68	98.52	74.81	91.22	92.29
M10	-	60.40	66.06	89.81	60.40	65.38	66.41

Table 6: Rejection frequencies (%) for Probit models (γ=2 and ρ=0.5)

Model	δ	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$	$\hat{S}_{B,\beta_2}$
$T = 100$							
M1,2,9	0	**5.84**	**6.60**	**5.19**	**5.85**	**4.65**	**-**
M1	0.4	34.36	30.74	51.53	34.36	52.92	-
	0.8	86.20	79.75	91.44	86.22	90.68	-
M2	0.4	17.53	11.24	16.73	17.53	13.43	-
	0.8	28.13	18.50	24.66	28.13	19.02	-
M3,4,10	0	**5.02**	**5.96**	**4.95**	**5.02**	**5.18**	**5.28**
M3	0.4	9.86	6.34	15.65	9.87	11.85	10.28
	0.8	23.00	15.57	36.65	23.00	23.79	25.13
M4	0.4	10.70	4.49	8.71	10.69	6.76	5.36
	0.8	23.22	10.59	28.23	23.21	17.95	17.94
$T = 500$							
M1,2,9	0	**5.41**	**6.63**	**5.66**	**5.38**	**5.20**	**-**
M1	0.4	95.36	95.02	94.47	95.35	97.55	-
	0.8	99.99	99.99	99.41	99.98	99.97	-
M2	0.4	38.50	31.30	43.24	38.51	37.64	-
	0.8	88.33	88.51	93.58	88.32	90.88	-
M3,4,10	0	**5.12**	**6.11**	**5.49**	**5.12**	**4.83**	**5.65**
M3	0.4	36.59	31.78	47.34	36.59	34.52	33.26
	0.8	91.04	88.13	95.91	91.04	86.00	86.28
M4	0.4	41.26	36.10	64.40	41.26	49.70	45.16
	0.8	81.11	94.54	99.12	81.11	94.00	94.34

Table 7: Rejection frequencies (%) for Logit models (γ=2 and ρ=0.5)

	δ	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$	$\hat{S}_{B,\beta_2}$
$T = 100$							
M1,2	0	**5.65**	**5.65**	**6.34**	**5.64**	**5.92**	-
M1	0.4	82.04	100	99.98	82.14	99.99	-
	0.8	99.75	100	95.81	99.79	94.94	-
M2	0.4	33.12	68.58	78.13	33.25	76.29	-
	0.8	88.87	99.08	99.15	88.92	99.32	-
M3,4	0	**5.79**	**5.42**	**5.53**	**5.79**	**5.49**	**5.86**
M3	0.4	90.57	87.72	99.13	90.78	91.24	83.58
	0.8	89.61	97.93	99.97	89.43	90.01	95.16
M4	0.4	62.13	95.97	96.90	62.25	85.51	86.63
	0.8	93.80	99.27	99.86	93.78	98.13	97.25
$T = 500$							
M1,2	0	**4.97**	**4.61**	**5.18**	**4.98**	**4.93**	-
M1	0.4	90.01	100	99.98	90.01	100	-
	0.8	89.42	100	98.35	89.42	96.65	-
M2	0.4	63.99	100	99.73	63.98	99.99	-
	0.8	100	99.98	99.79	100	100	-
M3,4	0	**4.85**	**4.64**	**5.10**	**4.85**	**5.12**	**4.94**
M3	0.4	95.54	100	99.99	95.69	100	100
	0.8	85.27	99.99	99.99	85.51	100	100
M4	0.4	92.73	100	100	92.77	100	100
	0.8	100	100	100	100	100	100

Table 8: Rejection frequencies (%) for Poisson count models (γ=2 and ρ=0.5)

	δ	$\hat{B}$	$\hat{S}_B$	$\hat{S}_{Bc}$	$\hat{S}_{B,\beta_0}$	$\hat{S}_{B,\beta_1}$	$\hat{S}_{B,\beta_2}$	$\hat{S}_{B,\alpha}$
$T = 100$								
M1,2	0	**5.38**	**9.87**	**6.86**	**6.47**	**6.51**	**-**	**7.97**
M1	0.4	15.29	92.96	84.60	13.96	86.26	-	67.52
	0.8	15.61	97.9	96.04	41.63	96.05	-	3.05
M2	0.4	11.87	21.51	28.36	5.75	35.49	-	3.98
	0.8	27.92	42.70	54.40	10.07	63.69	-	3.32
M3,4	0	**6.79**	**9.36**	**6.91**	**6.30**	**6.80**	**5.46**	**8.36**
	0.4	40.64	48.94	62.04	24.08	48.10	50.52	20.17
	0.8	29.98	87.43	88.63	37.40	67.30	72.22	36.44
M4	0.4	26.18	51.15	60.50	9.01	49.01	53.18	3.46
	0.8	44.81	67.65	79.94	16.28	64.83	71.34	2.97
$T = 500$								
M1,2	0	**4.97**	**6.16**	**4.79**	**5.26**	**4.81**	**-**	**5.58**
M1	0.4	42.42	99.50	96.69	11.74	94.46	-	65.46
	0.8	32.09	97.14	92.98	98.91	99.77	-	7.04
M2	0.4	27.60	90.16	93.25	12.55	96.81	-	5.84
	0.8	67.88	99.86	99.96	45.56	99.98	-	4.28
M3,4	0	**5.70**	**6.10**	**5.17**	**5.17**	**5.15**	**5.32**	**5.85**
M3	0.4	92.97	99.85	97.47	62.11	95.13	94.48	95.38
	0.8	35.31	97.66	95.02	68.77	91.39	93.75	42.22
M4	0.4	68.91	99.61	99.88	38.41	99.46	99.48	5.0
	0.8	94.38	100	100	67.47	99.99	99.99	4.67

Table 9: Rejection frequencies (%) for Negative Binomial 1 models (γ=2 and ρ=0.5)