Bartlett Identities Tests

Andrew Chesher [1], Geert Dhaene [2],
Christian Gouriéroux [3], and Olivier Scaillet [4]

June 1999

Abstract

In this note we propose a general testing procedure for parametric models based on Bartlett Identities. A well-known example is the Information Matrix test, which is based on the Bartlett Identity of order 1. The Identities are shown to induce a sequence of testable restrictions on the data generating process. When all the restrictions are considered jointly, they are often complete, in the sense that they are satisfied if and only if the model is correctly specified. We show that this is the case for normal, exponential and Poisson models. A test of the joint validity of an arbitrarily chosen subset of restrictions is proposed, and its first order asymptotic properties are presented. An interpretation of the test as a score test for neglected parameter heterogeneity is also given.

Résumé

Dans cette note, nous proposons une procédure de test générale de modèles paramétriques fondée sur des Identités de Bartlett. Un exemple classique de cette procédure est le test de la matrice d'information provenant de l'Identité de Bartlett d'ordre 1. Les Identités mènent à une suite de restrictions sur le processus générateur des données qui peuvent être testées. Lorsque toutes les restrictions sont considérées de manière jointe, elles sont souvent complètes, au sens où elles sont satisfaites si et seulement si le modèle est bien specifié. Nous le montrons pour le modèle normal, exponentiel et de Poisson. Un test de la validité jointe d'un sous-ensemble quelconque de restrictions est proposé, et ses propriétés asymptotiques de premier ordre sont examinées. Une interprétation du test comme test de score d'oubli d'hétérogénéité des paramètres est également donnée.

JEL classification: C12, C52.
Key words: Bartlett Identities, information matrix test, specification test.
Mots clés: Identités de Bartlett, test de la matrice d'information, test de spécification.

[1] University of Bristol, [2] Université de Mons-Hainaut and CORE, [3] CREST and CEPREMAP, [4] Université Catholique de Louvain. The first and fourth author gratefully acknowledge support under ESRC grant R008237386 and Belgian Program on Interuniversity Poles of Attraction (PAI nb. P4/01), respectively. We are grateful to Luc Bauwens for constructive comments. We have also benefited from discussions with M. Carrasco, J.-L. Prigent, and from numerous comments by seminar participants at Columbia University, Maastricht University, UQAM, CORE, CRDE and CREST.

1 Introduction

Many of the testing procedures used in practice are based on the likelihood principle. When null and alternative hypotheses are parametrically specified and nested these testing procedures include likelihood ratio, Wald and Lagrange multiplier tests (Breusch and Pagan (1980), Engle (1982)). To deal with non-nested hypotheses encompassing tests have been introduced (see Mizon and Richard (1986), Gouriéroux and Monfort (1994), Dhaene (1997), Dhaene, Gouriéroux and Scaillet (1998)).

But there also exist specification tests, developed directly from knowledge of a null hypothesis, and not explicitly oriented towards a specific alternative (Newey (1985), Tauchen (1985)). A well-known example is the information matrix (IM) test introduced by White (1982), which compares two estimators of the information matrix, for example minus the expected second order derivative of the log likelihood function and the sample variance of contributions to the associated score, unknown parameters being replaced by maximum likelihood estimates in both cases. It is known that this procedure is of a different nature and may be used to oppose the null hypothesis to a variety of alternatives, for instance to test for no skewness and no leptokurtosis of the error term in a Gaussian regression framework (White (1982)), or, in a wide class of models, to test for parameter constancy (Chesher (1984)). The different nature of such tests is also reflected in the number of conditions under test, $q-p$ in the classical tests, where q and p are the dimensions of maintained and null hypotheses respectively, but $p(p+1)/2$ (or some smaller number) for the IM test.

This paper is concerned with specification tests and explores the potential of the sequence of Bartlett Identities (Bartlett (1953a, 1953b), McCullagh (1987)) as a generator of useful specification tests (Chesher (1983)). Bartlett Identities (BIs) relate expectations of functions of derivatives of various orders of a log likelihood function. They imply moment restrictions (MRs) which will be satisfied when the likelihood function is well-specified. The complete sequence can imply MRs that are satisfied *only* if the likelihood function is correctly specified. The specification tests proposed here use sample analogues of these MRs. The IM test is an example of such a test because the information matrix identity, upon which it is based, is a Bartlett Identity. These BI based tests have several appealing features.

First, the test statistics are straightforward to compute. All elements of the statistics can be obtained from expressions for log likelihood derivatives whose derivation (numerical or analytic) can be automated, for example using symbolic computation tools. The resulting expressions are simple to calculate.

Second, when the restrictions at all (natural) orders are considered jointly, they are often complete, in the sense that they span the space of restrictions defined by the model on the data generating process. Then the MRs are jointly satisfied if and only if the model is correctly specified. Examples in which the set of MRs is complete are the normal, exponential and Poisson models.

Third, in some leading cases of interest the null distributions of the BI test statistics are exactly pivotal, that is their finite-sample distributions are parameter invariant. In these cases, as set out in HOROWITZ (1994) a Monte Carlo simulation provides arbitrarily precise estimates of exact critical values and the Monte Carlo tests described in DUFOUR and KIVIET (1996, 1998) and DUFOUR, FARHAT, GARDIOL and KHALAF (1998) deliver exact inference. This is highly desirable because analytic and Monte Carlo based evidence on the IM test (see e.g. CHESHER and SPADY (1991)) indicates that first order asymptotic approximations to its null distribution are frequently poor[1] and that size corrections based on second order approximations can be inaccurate.

In many other cases arising in econometric practice the BI tests are *partially* pivotal, that is their finite sample distributions are invariant with respect to a *subset* of parameters. Then the bootstrap described in a similar context in HOROWITZ (1994) delivers improved accuracy relative to first order asymptotic approximate inference and the Monte Carlo test procedure described in DUFOUR (1998) produces exact sized tests. Specifically the BI tests have this partially pivotal property in conditional models (i.e. given covariate values, X) in which some transformation of the response has conditional distribution given X which is in a location-scale family for fixed values of a subset of parameters with conditional location parameter $X\beta$. Index models of this sort are very common in applied parametric econometric analysis. This partial pivotal property occurs because (a) in such models all suitably scaled functions of all log likelihood derivative contributions of all orders evaluated at the MLE are partially pivotal (that is have distributions invariant with respect to β and the scale parameter as shown in CHESHER, DUMANGANE and SMITH (1998)) and (b) the BI test statistics are scale invariant functions of such contributions. In many cases the finite sample distributions of the tests conditional on X do depend upon X and then an ancillarity argument indicates that inference should be conducted conditional on the realised value of X, which is easily achieved using Monte Carlo or bootstrap tests.

The fourth appealing feature of the higher order BI tests is that they can deliver increased power relative to the BI test of order 1, that is the IM test. Power will be gained (respectively lost) if the induced higher order MRs do not hold (respectively hold). The examples given below show that the BI induced MRs often bear on the successive moments of the random variable considered. In these cases it is easy to see that the IM equality may hold when the likelihood function is misspecified and thus yield an IM test with low power, whereas higher order MRs may not hold and lead to more powerful tests. For example in the case of the exponential distribution the IM equality implies variance equal to squared mean, a relationship which can hold for many distributions other than the exponential.

In Section 2 the Bartlett Identities and associated moment restrictions are introduced in a general framework. We present a method to jointly test the empirical validity of a selection of BI associated MRs of arbitrary orders. In Section 3 we

[1] Monte-Carlo results available on request show that size distortions are even more pronounced for the BI test of order 2 than for the IM test, i.e. the BI test of order 1.

develop an interpretation of the complete sequence of BI tests as tests for parameter constancy. Some specific models in families with location or scale parameters are considered in Section 4. Section 5 deals with exponential family models and particular examples, namely normal, exponential, and Poisson models. As a simple example of conditional models, a normal regression model is considered in Section 6. Section 7 concludes.

2 Bartlett Identities Tests

2.1 Framework

We consider a sample of observations (y_t, x_t), $t = 1, \ldots, T$, with $y_t \in \mathbb{R}^k$ and $x_t \in \mathbb{R}^l$, independently drawn from a 'true' distribution, whose marginal and conditional distributions are denoted by P_X and $P_{Y|X}$. It is assumed that $P_{Y|X}$ has a probability density function (pdf) $f_0(y|x)$ w.r.t. a given measure $\bar{\mu}$ on $\mathbb{R}^k$ not depending on x. We consider a parametric model M for the conditional distribution of Y, given X. The model is defined by its conditional pdf :

$$M = \{f(y|x;\theta); \theta \in \Theta \subset \mathbb{R}^p\}.$$

The corresponding score function is given by :

$$s(y|x;\theta) = \frac{\partial}{\partial\theta} \log f(y|x;\theta).$$

The pseudo-true value of θ is :

$$\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmax}}\, E_X E_0 \left[\log f(Y|X;\theta)\right],$$

where E_X and E_0 are the mathematical expectations w.r.t. P_X and $P_{Y|X}$, respectively. We assume that θ_0 exists and is interior to the compact parameter space Θ. Further, we assume throughout that the true distribution and the model M allow us to take higher order derivatives of the log likelihood function and to interchange integration and differentiation. This framework also applies when Y has a discrete distribution if f is interpreted as a probability mass function.

2.2 Bartlett Identities and moment restrictions

By repeatedly differentiating the identity :

$$E_X E_\theta \left[s(Y|X;\theta)\right] = 0, \tag{1}$$

we immediately get the equalities defining the BIs.

Proposition 1 (Bartlett Identities) *For all θ interior to Θ :*

$$E_X E_\theta \left[r_n(Y|X;\theta)\right] = 0, \qquad n = 0, 1, \ldots, N,$$

where the functions r_n, $n = 0, 1, \ldots, N$, are recursively defined by :

$$\begin{aligned} r_0(Y|X;\theta) &= s(Y|X;\theta), \\ r_n(Y|X;\theta) &= (\text{vec } r_{n-1}(Y|X;\theta))s'(Y|X;\theta) \\ &\qquad + \frac{\partial}{\partial\theta'}\text{vec } r_{n-1}(Y|X;\theta), \qquad n = 1, 2, \ldots, N. \end{aligned}$$

For example the information matrix equality is obtained after one differentiation and corresponds to the BI, here referred to as of order 1 :

$$E_X E_\theta \left[s(Y|X;\theta)s'(Y|X;\theta)\right] + E_X E_\theta \left[\frac{\partial}{\partial\theta'}s(Y|X;\theta)\right] = 0. \tag{2}$$

The BIs have been used in the econometric literature for a variety of purposes. First, they have been found convenient from a numerical point of view since they avoid the need to compute expectations of higher derivatives of the log likelihood function (LANCASTER (1984)). If we need for instance to compute the expected value of the third derivative we may use the BI of order 2 :

$$-E_X E_\theta \left[s^{(2)}\right] = E_X E_\theta \left[\text{vec } \left(ss' + s^{(1)}\right) s'\right] + E_X E_\theta \left[\left[I \otimes s + s \otimes I\right] s^{(1)}\right],$$

omitting the arguments of the functions involved, and defining :

$$\begin{aligned} s^{(1)}(Y|X;\theta) &= \frac{\partial}{\partial\theta'}\text{vec } s(Y|X;\theta), \\ s^{(2)}(Y|X;\theta) &= \frac{\partial}{\partial\theta'}\text{vec } s^{(1)}(Y|X;\theta). \end{aligned}$$

Second, they help simplify higher order asymptotic expansions of the distributions of statistics (see e.g. BARTLETT (1953a, 1953b), HAYAKAWA (1977), HARRIS (1985)).

A third interesting use is found in specification testing. Indeed when model M is well-specified, i.e. $f_0(y|x) \in M$, the BI of order 1 (eq. (2)) induces the MR :

$$E_X E_0 \left[s(Y|X;\theta_0)s'(Y|X;\theta_0)\right] + E_X E_0 \left[\frac{\partial}{\partial\theta'}s(Y|X;\theta_0)\right] = 0, \tag{3}$$

which leads directly to the well-known IM test (WHITE (1982)). The relation (3) only takes into account the score associated with M, and is not *a priori* oriented towards a specific alternative. The MRs of arbitrary order follow directly from the BIs.

Corollary 1 (Moment Restrictions) *If M is well-specified, then :*

$$E_X E_0 \left[r_n(Y|X;\theta_0)\right] = 0, \qquad n = 0, 1, \ldots, N.$$

This corollary provides $N+1$ sets of MRs on the true distributions P_X and $P_{Y|X}$ that are automatically satisfied if M is well-specified. The number of MRs of order n equals p^{n+1}, of which at most $(n+p)!/\left((n+1)!(p-1)!\right)$ are independent. However the BI of order 0 does not entail a real MR because the condition

$$E_X E_0\left[s(Y|X;\theta_0)\right] = 0$$

is always satisfied since it serves to define the pseudo-true value, θ_0.

The test statistics introduced in Section 2.3 are based on the empirical counterparts of the MRs given in Corollary 1. Before studying inferential aspects of the statistics note that the moment conditions can be expressed in another form. Indeed equation (1) may be rewritten :

$$E_X E_\theta\left[\frac{\frac{\partial}{\partial\theta} f(Y|X;\theta)}{f(Y|X;\theta)}\right] = E_X \int \frac{\partial}{\partial\theta} f(Y|X;\theta)d\bar{\mu} = 0,$$

which leads after successive differentiations to :

$$E_X \int \frac{\partial^{n+1}}{\partial\theta_{k_1}\ldots\partial\theta_{k_{n+1}}} f(Y|X;\theta)d\bar{\mu} = 0, \qquad k_1 = 1,\ldots,p;\ldots;k_{n+1} = 1,\ldots,p.$$

Hence we deduce the following proposition which is essentially a transcription of Proposition 1.

Proposition 2 (Bartlett Identities) *For all θ interior to Θ :*

$$E_X E_\theta\left[\frac{\frac{\partial^{n+1}}{\partial\theta_{k_1}\ldots\partial\theta_{k_{n+1}}} f(Y|X;\theta)}{f(Y|X;\theta)}\right] = 0, \qquad n = 0,1,\ldots,N,$$

with : $k_1 = 1,\ldots,p;\ldots;k_{n+1} = 1,\ldots,p$.

Corollary 2 (Moment Restrictions) *If M is well-specified, then :*

$$E_X E_0\left[\frac{\frac{\partial^{n+1}}{\partial\theta_{k_1}\ldots\partial\theta_{k_{n+1}}} f(Y|X;\theta_0)}{f(Y|X;\theta_0)}\right] = 0, \qquad n = 0,1,\ldots,N,$$

with : $k_1 = 1,\ldots,p;\ldots;k_{n+1} = 1,\ldots,p$.

When the parameter is scalar ($p=1$), we get :

$$E_X E_0\left[\frac{\frac{\partial^{n+1}}{\partial\theta^{n+1}} f(Y|X;\theta_0)}{f(Y|X;\theta_0)}\right] = 0. \tag{4}$$

From this last expression it is easy to see that under mild conditions the set of moment restrictions characterize the true distribution.

Proposition 3 *If :*
(a) restriction (4) holds for some $\bar{\theta}$ *and* $n = 0, 1, \ldots,$
(b) for all θ*, the conditional parametric pdf admits the series expansion :*

$$f(y|x;\theta) = f(y|x;\bar{\theta}) + \sum_{n=1}^{\infty} \frac{\partial^n}{\partial\theta^n} f(y|x;\bar{\theta}) \frac{(\theta-\bar{\theta})^n}{n!}, \qquad \forall y, x, \tag{5}$$

(c) expectation and infinite sum can be interchanged :

$$E_X E_0 \left[\sum_{n=1}^{\infty} \frac{\frac{\partial^n}{\partial\theta^n} f(Y|X;\bar{\theta})}{f(Y|X;\bar{\theta})} \frac{(\theta-\bar{\theta})^n}{n!} \right] = \sum_{n=1}^{\infty} E_X E_0 \left[\frac{\frac{\partial^n}{\partial\theta^n} f(Y|X;\bar{\theta})}{f(Y|X;\bar{\theta})} \right] \frac{(\theta-\bar{\theta})^n}{n!},$$

(d) $L^1(P_{Y|X})$ *is generated by the likelihood ratios :* $\{f(Y|X;\theta)/f(Y|X;\bar{\theta}); \theta \text{ varying}\}$,

then for all y *and* x, $f(y|x;\bar{\theta}) = f_0(y|x)$.

Proof :

From (4) and (5) and condition (c)

$$E_X E_0 \left[\frac{f(Y|X;\theta) - f(Y|X;\bar{\theta})}{f(Y|X;\bar{\theta})} \right] = 0, \qquad \forall\theta.$$

Since $f(y|x;\theta)$ is a proper density function, for all θ and $\bar{\theta}$ and any $f_0(Y|X)$,

$$E_X E_0 \left[\frac{f(Y|X;\bar{\theta})}{f_0(Y|X)} \left(\frac{f(Y|X;\theta)}{f(Y|X;\bar{\theta})} - 1 \right) \right] = 0.$$

The last two expressions imply that for all θ :

$$E_X E_0 \left[\left(\frac{f(Y|X;\bar{\theta})}{f_0(Y|X)} - 1 \right) \left(\frac{f(Y|X;\theta)}{f(Y|X;\bar{\theta})} - 1 \right) \right] = 0$$

which, from (d), can only be true if $f(y|x;\bar{\theta}) = f_0(y|x)$ holds for all y and x. Q.E.D.

Finally, note that the BIs and associated MRs also hold conditionally on X. That is, the expectation operator E_X could be dropped everywhere. The BIs and the MRs presented here are a particular consequence of their conditional counterparts, obtained by integrating out X with respect to its true distribution P_X. Of course, other unconditional BIs and MRs could be obtained by integrating out X w.r.t. any other distribution, for example a distribution which is considered as a model for the unknown P_X. Often when working with models for conditional distributions P_X is left unspecified. Then one can proceed to consider tests of MRs that are obtained by integrating with respect to the true distribution. Consideration of ancillarity suggests that inference should be conducted conditional on the realised values of X, i.e. using the conditional MRs. As a practical matter this implies that when bootstrap or Monte Carlo BI tests are conducted the realised values of X should be held fixed across replications.

2.3 Test statistics

Given the sample of observations $(y_t, x_t), t = 1, \dots, T$, the pseudo-maximum likelihood (PML) estimator of θ is given by :

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\text{argmax}} \; T^{-1} \sum_{t=1}^{T} \log f(y_t | x_t; \theta).$$

The MRs for $n = 1, \dots, N$ given in Corollary 1 or 2 may be exploited to provide specification tests. As already mentioned the MR of order 0 cannot be used since its empirical counterpart corresponds to the first order condition for deriving the PML estimator, and is automatically satisfied. We define :

$$\hat{r}_n(\theta) = T^{-1} \sum_{t=1}^{T} r_n(y_t | x_t; \theta), \qquad n = 1, \dots, N,$$

and assume that the true distribution and the model M are sufficiently regular to ensure that, as $T \to \infty$, $\hat{\theta}_T$ converges a.s. to θ_0, and $\hat{r}_n(\theta)$ converges uniformly to $E_X E_\theta \left[r_n(Y|X;\theta) \right]$. Then, if M is well-specified, the statistics :

$$\hat{r}_n = \hat{r}_n(\hat{\theta}_T), \qquad n = 1, \dots, N,$$

all converge to zero, and specification tests of M can be based on this property. The asymptotic distribution of $\hat{r}_n$ under M is found from the expansion, up to order $o_p(1)$:

$$\begin{aligned} \sqrt{T} \text{ vec } \hat{r}_n &\simeq \sqrt{T} \text{ vec } \hat{r}_n(\theta_0) + \sqrt{T} \frac{\partial}{\partial \theta'} \text{vec } \hat{r}_n(\theta_0)(\hat{\theta}_T - \theta_0) \\ &\simeq \sqrt{T} \text{ vec } \hat{r}_n(\theta_0) + \left[E_X E_0 \frac{\partial}{\partial \theta'} \text{vec } r_n(Y \,|\, X; \theta_0) \right] \sqrt{T} \, (\hat{\theta}_T - \theta_0), \end{aligned}$$

and the standard expansion :

$$\sqrt{T} \, (\hat{\theta}_T - \theta_0) \simeq -B \frac{1}{\sqrt{T}} \sum_{t=1}^{T} s(y_t | x_t; \theta_0),$$

where $B = \left[E_X E_0 s^{(1)}(Y|X;\theta_0) \right]^{-1}$. It follows that, if M is well-specified,

$$\sqrt{T} \text{ vec } \hat{r}_n \xrightarrow{\text{d}} N(0, \text{Var} \, [Z_n]), \quad n = 1, \dots, N,$$

where :

$$Z_n = \text{vec } r_n(Y|X;\theta_0) - \left[E_X E_0 \frac{\partial}{\partial \theta'} \text{vec } r_n(Y \,|\, X; \theta_0) \right] B \, s(Y|X;\theta_0). \tag{6}$$

From this expression we deduce the joint distribution of the test statistics $\hat{r}_n$ and $\hat{r}_q$, with $n, q = 1, \dots, N$ and $n \neq q$, corresponding to the MRs of orders n and q.

Proposition 4 *If M is well-specified, then :*

$$\sqrt{T}\ \mathrm{vec}\ (\hat{r}_n, \hat{r}_q) \xrightarrow{\mathrm{d}} N(0, \Sigma_{n,q}), \quad n, q = 1, \ldots, N, \quad n \neq q,$$

where :

$$\Sigma_{n,q} = \begin{pmatrix} \mathrm{Var}\ [Z_n] & \mathrm{Cov}\ [Z_n, Z_q] \\ \mathrm{Cov}\ [Z_q, Z_n] & \mathrm{Var}\ [Z_q] \end{pmatrix},$$

and vec $(\hat{r}_n, \hat{r}_q)$ *is the stack of* vec $\hat{r}_n$ *and* vec $\hat{r}_q$.

The limiting covariance matrices can be estimated consistently by replacing pseudo-true values by maximum likelihood estimates. Mathematical expectations can be calculated analytically or can be replaced by empirical averages, giving rise to different estimates. For the construction of a test, it is natural to consider only a selection of the components of vec $(\hat{r}_n, \hat{r}_q)$. This selection has also to deal with the possible rank deficiency of $\Sigma_{n,q}$. Once this selection has been made and the limiting covariance matrix of the remaining components is non-singular, an asymptotic χ^2 testing procedure follows easily.

Let us now turn to the form involved in Corollary 2. We only study the one-dimensional case ($p = 1$) here. Define :

$$\hat{r}_n = T^{-1} \sum_{t=1}^{T} \frac{\dfrac{\partial^{n+1}}{\partial \theta^{n+1}} f(y_t | x_t; \hat{\theta}_T)}{f(y_t | x_t; \hat{\theta}_T)}, \qquad n = 1, \ldots, N.$$

When M is well-specified, the expansion is, up to order $o_p(1)$:

$$\begin{aligned} \sqrt{T}\hat{r}_n \quad \simeq \quad & T^{-1/2} \sum_{t=1}^{T} \frac{\dfrac{\partial^{n+1}}{\partial \theta^{n+1}} f(y_t | x_t; \theta_0)}{f(y_t | x_t; \theta_0)} \\ & - E_X E_0 \left[\frac{\dfrac{\partial^{n+1}}{\partial \theta^{n+1}} f(Y | X; \theta_0)}{f(Y | X; \theta_0)} \frac{\dfrac{\partial}{\partial \theta} f(Y | X; \theta_0)}{f(Y | X; \theta_0)} \right] \\ & \times \left(E_X E_0 \left[\frac{\dfrac{\partial}{\partial \theta} f(Y | X; \theta_0)}{f(Y | X; \theta_0)} \right]^2 \right)^{-1} T^{-1/2} \sum_{t=1}^{T} \frac{\dfrac{\partial}{\partial \theta} f(y_t | x_t; \theta_0)}{f(y_t | x_t; \theta_0)}. \end{aligned}$$

It is easy to see that $\sqrt{T}\hat{r}_n$ is equivalent to the residual mean in the regression of $(\partial^{n+1} f / \partial \theta^{n+1}) / f$ on $(\partial f / \partial \theta) / f$, and the asymptotic variance takes the form :

$$V_n = \mathrm{Var} \left[\frac{\dfrac{\partial^{n+1}}{\partial \theta^{n+1}} f(Y | X; \theta_0)}{f(Y | X; \theta_0)} \right] \left(1 - \mathrm{Corr}^{\ 2} \left[\frac{\dfrac{\partial}{\partial \theta} f(Y | X; \theta_0)}{f(Y | X; \theta_0)}, \frac{\dfrac{\partial^{n+1}}{\partial \theta^{n+1}} f(Y | X; \theta_0)}{f(Y | X; \theta_0)} \right] \right).$$

3 Heterogeneity

The BI test has a natural interpretation as a classical test against a class of alternatives under which parameters, which have been so far regarded as fixed, are subject to random variations. Suppose that the parameter θ possibly varies across observations with continuous, everywhere differentiable, distribution function $Q(\theta)$ so that $E[\theta] = \theta_0$. Let I be a p element vector of non-negative integers. Let J_s be the set of all distinct I such that $\sum_{j=1}^{p} I_j = L(I) = s$. Define the regular and absolute moments of θ :

$$\begin{aligned} \nu_I &= E[\prod_{j=1}^{p} (\theta_j - \theta_{0j})^{I_j}], \\ \bar{\nu}_I &= E[\prod_{j=1}^{p} |\theta_j - \theta_{0j}|^{I_j}], \end{aligned}$$

and assume that $\bar{\nu}_I$ is bounded for all $I \in J_{N+2}$. Let $\nabla_I f(y|x;\theta_0)$ denote the following partial derivative :

$$\nabla_I f(y|x;\theta_0) = \frac{\partial^{L(I)}}{\partial\theta_1^{I_1} \dots \partial\theta_p^{I_p}} f(y|x;\theta_0).$$

Assume all such derivatives are bounded in a neighbourhood of θ_0 for all $I \in J_{N+2}$.

Consider the distribution of Y given X marginal with respect to θ. Its density function is :

$$f(y|x) = \int f(y|x;\theta) dQ(\theta).$$

To derive the interpretation of BI test of order N we take a Taylor series expansion to order $N+1$ around $\theta = \theta_0$ and integrate with respect to θ. This gives the following approximation to $f(y|x)$:

$$f_N^a(y|x) = f(y|x;\theta_0) \left(1 + \sum_{i=2}^{N+1} \sum_{I \in J_i} \left(\prod_{j=1}^{p} (I_j!) \right)^{-1} \nu_I \frac{\nabla_I f(y|x;\theta_0)}{f(y|x;\theta_0)} \right). \tag{7}$$

If we write $\theta - \theta_0 = \tau U$ where $tr(\text{Var }[U]) = 1$ then $f(y|x) - f_N^a(y|x) = O(\tau^{N+2})$ as τ approaches zero, that is as random variation in θ becomes vanishingly small.

Consider a log likelihood function constructed using $f_N^a(y|x)$ in equation (7) as typical likelihood contributions[2]. This can be viewed as a local approximation to log likelihood functions generated under any random parameter model satisfying the restrictions imposed above. In the context of this approximate log likelihood

[2]Using the method employed in CHESHER and SMITH (1997) it is simple to re-express $f_N^a(y|x)$ as a proper density function without disturbing either the order of the approximation error or the argument of the reminder of this Section. This would lead to a Bartlett correctable likelihood ratio version of the BI tests.

function a score test of the hypothesis $H_0 : \nu_I = 0, \quad I \in \{J_2, \dots, J_{N+1}\}$ will consider departures from zero of sample averages :

$$T^{-1} \sum_{t=1}^{T} \frac{\nabla_I f(y_t|x_t; \hat{\theta}_T)}{f(y_t|x_t; \hat{\theta}_T)}$$

for $I \in \{J_2, \dots, J_{N+1}\}$, where $\hat{\theta}_T$ is the ML estimator in the fixed θ model (i.e. obtained under the null hypothesis). But these are precisely the statistics which form the basis of the BI tests. Therefore joint BI tests examining BIs of order up to N can be interpreted as score tests to detect random parameter variation, in a sense, refinements of the IM test for which this interpretation was provided in Chesher (1984).

4 Families with location or scale parameters

In location and scale families BI tests take simple forms as is shown below.

4.1 Families with a location parameter

Let us consider the model : $M = \{\varphi(y - \mu); \mu \in \mathbb{R}\}$, where φ is a given pdf and μ is an unknown location parameter. The BIs are :

$$E_\mu \left[(-1)^{n+1} \frac{\varphi^{(n+1)}(Y - \mu)}{\varphi(Y - \mu)} \right] = 0, \qquad n = 0, 1, \dots,$$

where $\varphi^{(n+1)}$ is the $n+1$-th derivative of φ.

Example : the normal distribution
For the Gaussian model the BIs are :

$$E_\mu \left[He_{n+1}(Y - \mu) \right] = 0, \qquad n = 0, 1, 2, \dots,$$

where He_n, $n = 0, 1, \dots$, are the Hermite polynomials (see Rodrigues' formula for orthogonal polynomials, Abramowitz and Stegun (1972), eq. 22.11.8 on p. 785) :

$$\begin{aligned} &He_0(y) = 1, \quad He_1(y) = y, \quad He_2(y) = y^2 - 1, \\ &He_3(y) = y(y^2 - 3), \quad He_4(y) = y^4 - 6y^2 + 3, \dots, \\ &He_{n+2}(y) = y He_{n+1}(y) - (n+1) He_n(y), \quad n \geq 0. \end{aligned} \tag{8}$$

Each additional MR deduced from these identities is effective and the MRs are jointly satisfied if and only if Y is normally distributed with unit variance.

Test statistics of the BI of order n for the normal model with known variance are of the form :

$$T_n = T \hat{r}_n^2 / \hat{V}_n,$$

where :

$$\hat{r}_n = T^{-1} \sum_{t=1}^{T} He_{n+1}(y_t - \hat{m}_1), \qquad \hat{m}_1 = T^{-1} \sum_{t=1}^{T} y_t,$$

and $\hat{V}_n$ is a consistent estimator of V_n (see (6)) :

$$V_n = \text{Var } \left[He_{n+1}(Y - m_1) - E_0[He'_{n+1}(Y - m_1)](Y - m_1)\right],$$

where He'_{n+1} is the derivative of He_{n+1} and $m_1 = E_0(Y)$. Since $E_0[He'_{n+1}(Y-m_1)] = 0$, and using the relation (see ABRAMOWITZ and STEGUN (1972), eqs. 22.1.2 and 22.2.15 on p. 775) :

$$\int_{-\infty}^{+\infty} \exp(-x^2/2)(He_n(x))^2 dx = \sqrt{2\pi} n!,$$

we deduce that :

$$V_n = (n+1)!.$$

Hence we may use as an estimator the exact asymptotic variance : $\hat{V}_{n,1} = V_n$. If mathematical expectations are not calculated analytically, we can also approximate V_n by :

$$\hat{V}_{n,2} = T^{-1} \sum_{t=1}^{T} \left[He_{n+1}(y_t - \hat{m}_1) - \left[T^{-1} \sum_{t=1}^{T} He'_{n+1}(y_t - \hat{m}_1)\right](y_t - \hat{m}_1)\right]^2.$$

If y_t, $t = 1, \ldots, T$, is an i.i.d. sample from a normal distribution with unit variance, then both $T_{n,1} = T\hat{r}_n^2/\hat{V}_{n,1}$ and $T_{n,2} = T\hat{r}_n^2/\hat{V}_{n,2}$ have an asymptotic $\chi^2(1)$ distribution.

Moreover, since the Hermite polynomials are orthogonal w.r.t. the weight function $\exp(-x^2/2)$, the statistics $\hat{r}_n$ and $\hat{r}_q$, $n \neq q$, are asymptotically uncorrelated. Therefore, if N is any finite set of ν positive integers, we get :

$$\sum_{n \in N} \frac{T\hat{r}_n^2}{(n+1)!} \xrightarrow{\text{d}} \chi^2(\nu).$$

4.2 Families with a scale parameter

Now consider the model : $M = \{\lambda\varphi(\lambda y); \lambda > 0\}$, where φ is a given pdf and λ is an unknown scale parameter. We obtain :

$$\begin{aligned}
r_0(y;\lambda) &= \frac{\varphi(\lambda y) + \lambda y \varphi^{(1)}(\lambda y)}{\lambda\varphi(\lambda y)}, \\
r_1(y;\lambda) &= \frac{2y\varphi^{(1)}(\lambda y) + \lambda y^2 \varphi^{(2)}(\lambda y)}{\lambda\varphi(\lambda y)}, \\
r_2(y;\lambda) &= \frac{3y^2\varphi^{(2)}(\lambda y) + \lambda y^3 \varphi^{(3)}(\lambda y)}{\lambda\varphi(\lambda y)},
\end{aligned}$$

and, by induction :

$$r_n(y;\lambda) = \frac{(n+1)y^n\varphi^{(n)}(\lambda y) + \lambda y^{n+1}\varphi^{(n+1)}(\lambda y)}{\lambda\varphi(\lambda y)}.$$

The BIs are therefore:

$$E_\lambda\left[\frac{(n+1)Y^n}{\lambda}\frac{\varphi^{(n)}(\lambda Y)}{\varphi(\lambda Y)} + Y^{n+1}\frac{\varphi^{(n+1)}(\lambda Y)}{\varphi(\lambda Y)}\right] = 0, \qquad n = 0, 1, \ldots.$$

Example : the exponential distribution
Consider the exponential model : $M = \{\lambda\exp(-\lambda y); y \geq 0, \lambda > 0\}$. We get :

$$\begin{aligned} r_0(y;\lambda) &= \frac{1}{\lambda} - y, \\ r_1(y;\lambda) &= -y(\frac{2}{\lambda} - y), \\ r_2(y;\lambda) &= y^2(\frac{3}{\lambda} - y), \end{aligned}$$

and :

$$r_n(y;\lambda) = (-y)^n(\frac{n+1}{\lambda} - y).$$

Taking expectations with respect to the true distribution of Y gives :

$$E_0[r_0(Y;\lambda_0)] = \frac{1}{\lambda_0} - E_0[Y] = 0,$$

thereby defining λ_0, and the MRs are :

$$\begin{aligned} E_0[r_1(Y;\lambda_0)] &= -2m_1^2 + m_2 = 0, \\ E_0[r_2(Y;\lambda_0)] &= 3m_1m_2 - m_3 = 0, \\ &\vdots \\ E_0[r_n(Y;\lambda_0)] &= (-1)^n((n+1)m_1m_n - m_{n+1}) = 0, \end{aligned}$$

where $m_i = E_0[Y^i]$. The MR of order n is an effective constraint on the underlying distribution, in addition to the MRs of lower order. Furthermore, all the MRs for $n = 1, 2, \ldots$, are simultaneously satisfied if and only if Y is exponentially distributed. Indeed, we have $m_1 = 1/\lambda_0$, and from the MRs we get :

$$m_{n+1} = (n+1)m_1m_n, \qquad n = 1, 2, \ldots,$$

which holds if and only if $m_n = n!/\lambda_0^n$, and we deduce a unique expression for the characteristic function.

Test statistics of the BI of order n for the exponential model are of the form :

$$T_n = T\hat{r}_n^2/\hat{V}_n,$$

where :

$$\begin{aligned}\hat{r}_n &= (-1)^n \left((n+1)\hat{m}_1\hat{m}_n - \hat{m}_{n+1}\right), \\ \hat{m}_i &= T^{-1}\sum_{t=1}^{T} y_t^i, \qquad i = 1, 2, \ldots,\end{aligned}$$

and $\hat{V}_n$ is a consistent estimator of V_n :

$$\begin{aligned}V_n &= \text{Var}\left[(-Y)^n((n+1)m_1 - Y) - (-1)^{(n+1)}(n+1)m_1^2 m_n \left(-\frac{1}{m_1} + \frac{Y}{m_1^2}\right)\right] \\ &= \text{Var}\left[Y^{n+1} - (n+1)m_1 Y^n - (n+1)m_n Y + (n+1)m_1 m_n\right], \qquad (9)\end{aligned}$$

where $m_i = E_0[Y^i] = i! m_1^i$. As noted earlier, there are a variety of estimators of V_n. Calculating the variance of the bracketed term in (9) analytically in terms of $\lambda_0 = m_1^{-1}$ and then substituting $\hat{\lambda} = \hat{m}_1^{-1}$ for λ_0 yields the following estimator of V_n :

$$\hat{V}_{n,1} = (n+1)^2[(2n)! - (n!)^2]\hat{m}_1^{2n+2}.$$

Noting that the bracketed term in (9) has zero expectation, an alternative estimator is found as

$$\begin{aligned}\hat{V}_{n,2} &= T^{-1}\sum_{t=1}^{T}\left[y_t^{n+1} - (n+1)\hat{m}_1 y_t^n - (n+1)\hat{m}_n y_t + (n+1)\hat{m}_1\hat{m}_n\right]^2 \\ &= \hat{m}_{2n+2} - 2(n+1)\hat{m}_1\hat{m}_{2n+1} + (n+1)^2\hat{m}_1^2\hat{m}_{2n} - 3(n+1)^2\hat{m}_1^2\hat{m}_n^2 \\ &\quad +2(n+1)(n+2)\hat{m}_1\hat{m}_n\hat{m}_{n+1} + (n+1)^2\hat{m}_2\hat{m}_n^2 - 2(n+1)\hat{m}_n\hat{m}_{n+2}.\end{aligned}$$

If y_t, $t = 1, \ldots, T$, is an i.i.d. sample from an exponential distribution, then both $T_{n,1} = T\hat{r}_n^2/\hat{V}_{n,1}$ and $T_{n,2} = T\hat{r}_n^2/\hat{V}_{n,2}$ have an asymptotic $\chi^2(1)$ distribution.

To estimate the asymptotic covariance between $\sqrt{T}\hat{r}_n$ and $\sqrt{T}\hat{r}_q$, we may use either empirical averages or :

$$\hat{\Sigma}_{n,q} = (-1)^{n+q}(n+1)(q+1)[(n+q)! - n!q!]\hat{m}_1^{n+q+2}.$$

Example : the normal distribution

For the Gaussian model with known mean, assumed to be zero without loss of generality, the BIs are :

$$E_\lambda\left[-\frac{(n+1)Y^n}{\lambda}He_n(\lambda Y) + Y^{n+1}He_{n+1}(\lambda Y)\right] = 0, \qquad n = 0, 1, 2, \ldots. \qquad (10)$$

Using the recurrence relation (8) between Hermite polynomials they can be rewritten as :

$$E_\lambda\left[Y^n He_{n+2}(\lambda Y)\right] = 0, \qquad n = 0, 1, 2, \ldots.$$

From the explicit expression of He_n (ABRAMOWITZ and STEGUN (1972), eq. 22.3.11 on p. 775) :

$$He_n(x) = n! \sum_{k=0}^{[n/2]} \frac{(-1)^k x^{n-2k}}{k! 2^k (n-2k)!},$$

we see that the BI of order n bears only on the even moments of Y up to order $2n+2$. Furthermore it is obvious that $\lambda_0^2 = 1/E_0[Y^2]$, and it can be checked that Y satisfies all MRs if and only if $E_0[Y^{2(n+1)}] = (\lambda_0)^{-2(n+1)}(2(n+1))!/(2^{n+1}(n+1)!)$, $n = 0, 1, 2, \ldots$, which characterizes the even moments of the normal distribution. Observe that the test is incomplete in the sense that the MRs induced by the BIs do not span the space of all restrictions induced by the model. It only allows to check for the distributional properties of the minimal sufficient transform Y^2.

5 Exponential families

Now consider the one parameter exponential family :

$$f(y;\mu) = \exp(A(\mu) + B(y) + C(\mu)y).$$

We have :

$$\frac{\partial^n f(y;\mu)}{\partial \mu^n} = f(y;\mu) h_n(y;\mu),$$

where $h_n(y;\mu)$ satisfies the recurrence equation :

$$\begin{aligned} h_1(y;\mu) &= A^{(1)}(\mu) + C^{(1)}(\mu)y, \\ h_n(y;\mu) &= (A^{(1)}(\mu) + C^{(1)}(\mu)y) h_{n-1}(y;\mu) + h_{n-1}^{(1)}(y;\mu), \end{aligned}$$

with $A^{(1)}(\mu) = \partial A(\mu)/\partial\mu$, $C^{(1)}(\mu) = \partial C(\mu)/\partial\mu$, and $h_{n-1}^{(1)}(\mu) = \partial h_{n-1}(y;\mu)/\partial\mu$. The BIs are thus equal to :

$$E_\mu \left[h_{n+1}(Y;\mu) \right] = 0, \qquad n = 0, 1, \ldots.$$

Let us note that $h_n(Y;\mu)$ is a polynomial of degree n as soon as $C^{(1)}(\mu)$ is non zero for all μ, which is an identifiability condition on the parameter.

Example : the Poisson distribution
The model is $M : \{f(y;\mu) = e^{-\mu}\mu^y/y!; y \in \mathbb{N}, \mu > 0\}$. By differentiating w.r.t. to the parameter μ we deduce :

$$\begin{aligned} \frac{1}{f(y;\mu)} \frac{\partial^n f(y;\mu)}{\partial\mu^n} &= \frac{e^\mu}{\mu^y} \sum_{j=0}^{n} C_n^j \frac{\partial^{n-j}}{\partial\mu^{n-j}}(e^{-\mu}) \frac{\partial^j}{\partial\mu^j}(\mu^y) \\ &= \sum_{j=0}^{\min\{y,n\}} (-1)^{n-j} C_n^j \mu^{-j} \frac{y!}{(y-j)!}. \end{aligned}$$

The factorial moments of order $0, 1, \ldots, n$, enter successively into the various BIs. For example, the BI of order 2 will provide the standard condition which relates the mean and the variance of a Poisson distribution and is the basis of the overdispersion test (see e.g. GURMU (1991)). When we take the expectation w.r.t. the true distribution of Y, the MR of order n will lead to a restriction bearing only on the factorial moment of order n if all MRs of lower order are satisfied. By using the expression for the factorial moment of the Poisson distribution (JOHNSON, KOTZ and KEMP (1992) p. 156) and the link between the factorial moment generating function and the characteristic function for a discrete variable, we deduce that all MRs are satisfied if and only if Y follows a Poisson distribution. Applying the general procedure outlined in Section 2 will thus lead to a test of the Poisson model.

6 The normal regression model

Consider the normal regression model

$$Y \mid X \sim N(X\theta, 1),$$

with one regressor X and fixed variance not depending on X. The BIs are found along the same lines as in Section 4.1. We get :

$$E_X E_\theta[X^{n+1} He_{n+1}(Y - X\theta)] = 0, \qquad n = 0, 1, \ldots.$$

The corresponding MRs are :

$$E_X E_0[X^{n+1} He_{n+1}(Y - X\theta_0)] = 0, \qquad n = 0, 1, \ldots,$$

where θ_0 is defined by $E_X E_0[X(Y - X\theta_0)] = 0$. Thus, the MRs correspond to orthogonality conditions between powers of X and the Hermite polynomials in the 'pseudo-true disturbance' $Y - X\theta_0$. The test statistic of the BI of order n is $T_n = T\hat{r}_n^2/\hat{V}_n$, where now :

$$\hat{r}_n = T^{-1} \sum_{t=1}^{T} x_t^{n+1} He_{n+1}(\hat{u}_t),$$

with $\hat{u}_t$ the least squares residual in the regression of y_t on x_t $(t = 1, \ldots, T)$. The covariance matrix V_n is :

$$V_n = \text{Var}\left[X^{n+1} He_{n+1}(Y - X\theta_0)\right] = (n+1)!\, E_X(X^{2n+2}),$$

and is consistently estimated by :

$$\hat{V}_{n,1} = (n+1)!\, \hat{m}_{2n+2}(X),$$

where $\hat{m}_j(X)$ is the j-th sample moment of X. Again, due to the orthogonality of Hermite polynomials, the statistics $\hat{r}_n$ and $\hat{r}_q$, $n \neq q$, are asymptotically uncorrelated, and we get :

$$\sum_{n \in N} \frac{T\hat{r}_n^2}{(n+1)!\, \hat{m}_{2n+2}(X)} \xrightarrow{\text{d}} \chi^2(\nu),$$

for any finite set N of ν positive integers.

The previous statistics may also be used in a time series framework where X is the lagged value of Y. In this case they complete naturally the standard Box-Pierce (Portmanteau) statistics (see GOURIÉROUX and JASIAK (1999)). Indeed the latter is appropriate for checking the absence of linear time dependence of any order, whereas the BI based tests are appropriate for checking the absence of any nonlinear time dependence of order one. Furthermore, following the interpretation given in Section 3, they can also be viewed as tests for fixed versus random coefficient autoregressions.

7 Conclusion

In this paper, we have derived a generalization of the IM test based on the Bartlett Identities. These identities impose moment restrictions on the observed data if the hypothesized likelihood is correctly specified and the restrictions are easy to test. Of particular interest is the finding that the specification testing procedure delivers in principle complete tests of normal, exponential and Poisson models. Since the tests can involve high order sample moments first order asymptotic approximations may be poor but bootstrap and Monte Carlo test methods will deliver accurate inference in many situations arising in practice.

References

[1] Abramowitz, M. and I. Stegun (1972), *Handbook of Mathematical Functions*, Dover Publications, New York.

[2] Bartlett, M.S. (1953a), "Approximate Confidence Intervals", *Biometrika*, 40, 21-19.

[3] Bartlett, M.S. (1953b), "Approximate Confidence Intervals II. More Than One Parameter", *Biometrika*, 40, 306-317.

[4] Breusch, T. and A. Pagan (1980), "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics", *Review of Economic Studies*, **47**, 238–253.

[5] Chesher, A. (1983), "The Information Matrix Test, Extensions and Omnibus Tests of Specification", Discussion Paper 98, Centre for Econometrics and Decision Sciences, University of Florida.

[6] Chesher, A. (1984), "Testing for Neglected Heterogeneity", *Econometrica*, **52**, 865–872.

[7] Chesher, A., Dumangane, M. and R. Smith (1998), "Duration Response Measurement Error", University of Bristol Department of Economics Discussion Paper No. 98/451.

[8] Chesher, A. and R. Spady (1991), "Asymptotic Expansions of the Information Matrix Test Statistic", *Econometrica*, **59**, 787–815.

[9] Chesher, A. and R. Smith (1997), "Likelihood Ratio Specification Tests", *Econometrica*, **65**, 627–646.

[10] Dhaene, G. (1997), *Encompassing. Formulation, Properties and Testing*, Lecture Notes in Economic and Mathematical Systems, n^o 446, Springer-Verlag, Berlin.

[11] Dhaene, G., Gouriéroux, C. and O. Scaillet (1998), "Instrumental Models and Indirect Encompassing", *Econometrica*, **66**, 673–688.

[12] Dufour, J.-M. (1998), "Monte Carlo Tests with Nuisance Parameters : A General Approach to Finite-Sample Inference and Nonstandard Asymptotics", DP CRDE.

[13] Dufour, J.-M., Farhat, A., Gardiol, L. and L. Khalaf (1998), "Simulation-Based Finite Sample Normality Tests in Linear Regressions", *Econometrics Journal*, **1**, 154–173.

[14] Dufour, J.-M. and J. Kiviet (1996), "Exact Tests for Structural Change in First-Order Dynamic Models ", *Journal of Econometrics*, **70**, 39–68.

[15] Dufour, J.-M. and J. Kiviet (1998), "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models", *Econometrica*, **66**, 79–104.

[16] Engle, R. (1982), "A General Approach to Lagrange Multiplier Model Diagnostics", *Journal of Econometrics*, **20**, 83–104.

[17] Gouriéroux, C. and J. Jasiak (1999), "Nonlinear Innovations and Impulse Response Functions", mimeo.

[18] Gouriéroux, C. and A. Monfort (1994), "Testing Non-Nested Hypotheses", in *Handbook of Econometrics*, **IV**, ed. by R. Engle and D. McFadden, Elsevier Science, North-Holland, Amsterdam, 2583–2637.

[19] Gouriéroux, C. and A. Monfort (1995), *Statistics and Econometric Models*, Vol. 2, Cambridge University Press, Cambridge.

[20] Gurmu, S. (1991), "Tests for Detecting Overdispersion in the Positive Poisson Regression Model", *Journal of Business and Economic Statitics*, **9**, 215–222.

[21] Harris, P. (1985), "An Asymptotic Expansion for the Null Distribution of the Efficient Score Statistic", *Biometrika*, **72**, 653–659.

[22] Hayakawa, T. (1977), "The Likelihood Ratio Criterion and the Asymptotic Expansion of its Distribution", *Annals of the Institute of Statistical Mathematics*, **29**, 359–378.

[23] Horowitz, J. (1994), "Bootstrap-based Critical Values for the Information Matrix Test", *Journal of Econometrics*, **61**, 395–411.

[24] Johnson, N., Kotz, S. and N. Balakrishnan (1994), *Continuous Univariate Distributions*, Vol. 1, Wiley, New York.

[25] Johnson, N. Kotz, S. and A. Kemp (1992), *Univariate Discrete Distributions*, Wiley, New York.

[26] Lancaster, T. (1984), "The Covariance Matrix of the Information Matrix Test", *Econometrica*, **52**, 1051–1054.

[27] McCullagh, P. (1987), *Tensor Methods in Statistics*, Chapman Hall, London.

[28] Mizon, G. and J.-F. Richard (1986), "The Encompassing Principle and its Applications to Testing Non-nested Hypotheses", *Econometrica*, **54**, 657–678.

[29] Newey, W. (1985), "Maximum Lihelihood Specification Testing and Conditional Moment Tests", *Econometrica*, **53**, 1047–1070.

[30] Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Lihelihood Models", *Journal of Econometrics*, **30**, 415–443.

[31] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, **50**, 1–26.