THE INFORMATION MATRIX TEST AND ITS IMPLIED ALTERNATIVE HYPOTHESES

A. Colin Cameron
and
Pravin K. Trivedi

Working Paper Series
#372

Department of Economics
University of California
Davis, California 95616-8578

THE INFORMATION MATRIX TEST AND ITS IMPLIED ALTERNATIVE HYPOTHESES

by

A. Colin Cameron
Department of Economics
University of California
Davis, California 95616-8578

Pravin K. Trivedi
Department of Economics
Indiana University
Bloomington, Indiana 47405

Working Paper Series No. 372
December 1990

ABSTRACT

The information matrix (IM) test of White (1982) is a model specification test obtained by specifying a null hypothesis model only. A criticism often made is that failure to specify an alternative hypothesis model makes it difficult, in a general setting, to interpret what types of departure the IM test is testing against. In this paper it is shown how the IM test can be interpreted as a test against an alternative hypothesis.

Some Key Words: information matrix tests; conditional moment specification tests; score tests; random parameter heterogeneity; generalized linear models; linear exponential family; heteroskedasticity; symmetry; kurtosis.

Acknowledgements: This paper has benefited from presentation of a related paper at University of California - Los Angeles and Davis, and from discussion with Dick Jefferis.

1. INTRODUCTION

The information matrix (IM) test of White (1982) is an intuitively appealing model specification test, and is easily implemented using results of Chesher (1983) and Lancaster (1984). Yet it is not widely adopted as a model specification test, in part because of lack of knowledge as to what alternative hypothesis the null hypothesis model is being tested against.

In this paper we provide a general interpretation of the IM test as a test that a certain conditional moment has expectation zero under the null hypothesis and expectation equal to a specific functional form under an alternative hypothesis.

This interpretation of the IM test is more general than that in other studies in specific settings, such as the linear regression model with constant variance normally distributed error, discussed in White (1982, p.12) and analyzed in detail by Hall (1987), and the linear regression model with autocorrelated errors, analyzed by Bera and Lee (1990). At the same time it provides a more tightly specified alternative hypothesis than does the general treatment of Chesher (1984).

The interpretation of the IM test as being a test against an alternative hypothesis is possible because parametric econometric models with densities depending on q parameters, say, are typically based on densities depending on underlying parameters of dimension much less than q. For example, for the linear regression model under normality, the underlying parameters are the mean and variance. These two underlying parameters are in turn modeled to depend on explanatory variables and q unknown parameters.

The particular alternative hypothesis obtained can be interpreted as arising from local parameter heterogeneity in the underlying parameters, with a specific functional form for the covariance matrix of the underlying parameters. The IM test is accordingly a score test against a quite specific form of heterogeneity in the underlying parameters. By contrast, Chesher (1984) considered heterogeneity in all q parameters, in which case the IM test is a score test against quite general forms of heterogeneity in the q parameters.

The general theory is presented in section 2. Substantive examples of IM tests for commonly-used regression models based on the linear exponential family (one underlying parameter) and for the regression model with normally distributed homoscedastic error (two underlying parameters) are given in section 3.

Previous in-depth studies of IM tests for specific examples have implicitly restricted analysis, by considering models that are linear in parameters. More generally, the underlying parameters may be a non-linear function of the q parameters, e.g. a non-linear regression model. Then the results of section 2 hold, provided the derivative of the log density with respect to the underlying parameters has expectation zero. Results when this condition is instead not imposed are given in section 4. Section 5 concludes.

2. General Theory

2.1 Information Matrix Tests

We are interested in explaining dependent variables, a vector y_t, conditional on pre-determined explanatory variables, a vector X_t. Statistical inference is based on an assumed parameterized density function $f(y_t | X_t,\theta)$, denoted $f(y_t,X_t,\theta)$, where θ is a $q \times 1$ parameter vector, that satisfies the regularity conditions of White (1982). This paper focuses on cross-section data, $\{(y_t, X_t), t = 1,\ldots,T\}$, independent across t.

The information matrix equality implies that $\mathbb{E}_0[D(y_t,X_t,\theta) \mid X_t] = 0$, where the subscript 0 denotes expectation with respect to the assumed density $f(y_t,X_t,\theta)$, and

$$(2.1) \qquad D(y_t,X_t,\theta) = \nabla_\theta^2 L(y_t,X_t,\theta) + \nabla_\theta L(y_t,X_t,\theta)\cdot(\nabla_\theta L(y_t,X_t,\theta))' \quad ,$$

where $L(y_t,X_t,\theta) = \log(f(y_t,X_t,\theta))$, and ∇_θ^j denotes the j-th derivative with respect to θ. Define

$$(2.2) \qquad d(y_t,X_t,\theta) = \mathrm{vech}(D(y_t,X_t,\theta)) \quad ,$$

where $\mathrm{vech}(\cdot)$ is the "vector half" operator which stacks the lower triangular part of a symmetric matrix into a column vector. Thus $d(y_t,X_t,\theta)$ contains the $q(q+1)/2$ unique elements of the $q\mathrm{x}q$ symmetric matrix $D(y_t,X_t,\theta)$.

The information matrix (IM) test of White (1982) is based on the $q(q+1)/2$ moment conditions[1]:

$$(2.3) \qquad \mathbb{E}_0[d(y_t,X_t,\theta) \mid X_t] = 0 \quad .$$

IM tests are statistical tests of the departure from zero of the corresponding sample moment, $d_T(\hat{\theta}_T)$, where

$$(2.4) \qquad d_T(\theta) = T^{-1} \sum_{t=1}^{T} d(y_t,X_t,\theta) \quad ,$$

and $\hat{\theta}_T$ is an estimator consistent for θ under the true model.[2]

We note that the IM Test is a special case of the conditional moment (CM) tests of Newey (1985) and Tauchen (1985). In their more general framework,

$d(y_t, X_t, \theta)$ in (2.3) and (2.4) may be any function with expectation zero under the assumed model, not just that function defined by (2.1). The CM test framework is used below.

2.2 An Underlying Alternative for the IM Test

For the case where y_t is i.i.d., there is a large menu of density functions of the form $f(y_t, \eta)$, where η is a hx1 vector. In regression analysis, the dependence of y_t on explanatory variables X_t is captured by replacing η_i by $\eta_{it} = \eta_i(X_t, \theta_i)$, $i = 1,\ldots,h$, where θ_i is q_ix1, and $\theta = (\theta_1', \ldots, \theta_h')'$ is the qx1 vector of parameters in section 2.1. The vector of underlying parameters η has dimension (h) that is considerably less than q. For example, in the classical linear regression model under normality, y_t is $N(\eta_{1t}, \eta_{2t})$, where $\eta_{1t} = X_t'\beta$ and $\eta_{2t} = \sigma^2$, so $h = 2$ and $q = 1 + \dim(\beta)$.

The assumed density is therefore of the form:

$$f(y_t, X_t, \theta) = f(y_t, \eta_1(X_t, \theta_1), \ldots, \eta_h(X_t, \theta_h)) \quad . \tag{2.5}$$

Given this representation of the density, the moment conditions tested by the IM test are obtained by application of the chain rule of differentiation.

Proposition 1: For the density (2.5), the (partial) IM test is a test of the q(q+1)/2 unique moment restrictions:

$$\mathbb{E}_0[g_{ij}(X_t, \theta) \cdot H_{ij}(y_t, X_t, \theta) \mid X_t] = 0, \quad i = 1,\ldots,j, \quad j = 1,\ldots,h, \tag{2.6}$$

where $H_{ij}(y_t, X_t, \theta) \equiv H_{ij}(y_t, \eta(X_t, \theta))$ is a scalar defined by

(2.7) $H_{ij}(y_t,X_t,\theta)$

$$= \nabla^2_{\eta_i\eta_j} L(y_t,\eta(X_t,\theta)) \quad + \quad \nabla_{\eta_i} L(y_t,\eta(X_t,\theta))\cdot\nabla_{\eta_j} L(y_t,\eta(X_t,\theta)),$$

$L(y_t,\eta(X_t,\theta)) \;=\; \log(f(y_t,\eta(X_t,\theta)))$, $\nabla_{\eta_i} \equiv \partial/\partial\eta_i$, $\nabla^2_{\eta_i\eta_j} \equiv \partial^2/\partial\eta_i\partial\eta_j'$, and

$$\text{(2.8)} \qquad g_{ij}(X_t,\theta) \;=\; \mathrm{vec}(\nabla_{\theta_i}\eta_i(X_t,\theta_i)\cdot(\nabla_{\theta_j}\eta_j(X_t,\theta_j))'), \qquad i \neq j,$$
$$= \mathrm{vech}(\nabla_{\theta_i}\eta_i(X_t,\theta_i)\cdot(\nabla_{\theta_i}\eta_i(X_t,\theta_i))'), \qquad i = j,$$

where $\nabla_{\theta_i} \equiv \partial/\partial\theta_i$, vec(•) denotes vectorization, and vech(•) denotes vectorization choosing unique elements. The vector g_{ij} is of dimensions q_iq_j, $i \neq j$, and $q_i(q_i+1)/2$, $i = j$. ■

For derivation of Proposition 1, see the Appendix. The term "partial" IM test is used in the sense that the second term in (A.3) is neglected. This second term actually equals zero when $\nabla^2_\theta\eta_i(X_t,\theta) = 0$, $i = 1,\ldots,h$, which is the case in existing detailed studies of applications of the IM test, notably Hall (1987), Orme (1990), Bera and Lee (1990). More generally, however, $\nabla^2_\theta\eta_i(X_t,\theta) \neq 0$, in which case the discussion in sections 2 and 3 needs to be viewed as an analysis conditional on the assumption $E_0[\nabla_\eta L(y_t,\eta(X_t,\theta)) \mid X_t] = 0$, which also permits neglect of the second term. The more cumbersome analysis including the second term is deferred to section 4.

For given i and j, a test of the moment conditions (2.6) can be viewed, following Newey (1985), as being based on the underlying scalar moment condition:

$$\text{(2.9)} \qquad H_0\colon\ \mathbb{E}_0[H_{ij}(y_t,X_t,\theta) \mid X_t] \;=\; 0 \quad ,$$

since (2.6) implies (2.9) by the law of iterated expectations. i.e. IM tests based on the moment conditions (2.6) are tests of the departure of (2.9) in the direction in which the vector function $\mathrm{vec}(\nabla_{\theta_i}\eta_i(X_t,\theta_i)\cdot(\nabla_{\theta_j}\eta_j(X_t,\theta_j))')$ of the explanatory variables is correlated with $H_{ij}(y_t,X_t,\theta)$.

Thus the IM test of q(q+1)/2 moment conditions, seemingly with no alternative, is in fact a test of h(h+1)/2 underlying moment conditions (2.9). The departure of each individual underlying moment condition is tested in a specific direction, given by the orthogonality condition (2.6).

A more precise statement of the direction in which departure is being tested is the following.

Corollary: IM Tests based on (2.6) are optimal regression-based CM tests of

(2.9) $$H_0\colon\ \mathbb{E}_0[H_{ij}(y_t,X_t,\theta)\mid X_t] = 0\ ,$$

against the alternative:

(2.10) $$H_1\colon\ \mathbb{E}_1[H_{ij}(y_t,X_t,\theta)\mid X_t] = \Sigma_{ij}(X_t,\theta)\cdot g_{ij}(X_t,\theta)'\cdot\alpha_{ij}\ ,$$

where the subscript 1 denotes expectation with respect to the true d.g.p.,

(2.11) $$\Sigma_{ij}(X_t,\theta) = \mathrm{Var}_0(H_{ij}(y_t,X_t,\theta)),$$

is a positive definite matrix, and α_{ij} is a px1 vector of additional unknown parameters. ■

Optimal regression-based CM tests are presented in Cameron and Trivedi (1990). The term "regression-based" CM tests, henceforth RBCM tests, is used in the following sense. Tests of the moment condition (2.9) under H_0 against (2.10) under H_1 are tests of whether $\alpha_{ij} = 0$. The obvious basis for such a test is the estimated coefficient of α_{ij} in the regression corresponding to (2.10):

$$(2.12) \qquad H_{ij}(y_t,X_t,\theta) = \Sigma_{ij}(X_t,\theta)\cdot g_{ij}(X_t,\theta)'\cdot\alpha_{ij} + \varepsilon_{ij,t},$$

where $\varepsilon_{ij,t}$ is a heteroskedastic error under H_0, with variance given in (2.11).

The term "optimal" RBCM test is used to indicate that the most powerful test of $\alpha_{ij} = 0$ against local alternatives $\alpha_{ij} = T^{-1/2}\gamma_{ij}$, based on the regression (2.12) will use the weighted least squares estimator, which divides terms by $\Sigma_{ij}(X_t,\theta)^{1/2}$ and equals:

$$(2.13) \quad \hat{\alpha}_{ij} = \{\sum_{t=1}^{T} g_{ij}(X_t,\theta)\cdot\Sigma_{ij}^{-1}(X_t,\theta)\cdot g_{ij}(X_t,\theta)'\}^{-1}\cdot\sum_{t=1}^{T} g_{ij}(X_t,\theta)\cdot H_{ij}(y_t,X_t,\theta)$$

Tests based on $\hat{\alpha}_{ij}$ are equivalent to tests based on the second summation term only, but this is just the CM test based on the sample moment corresponding to (2.6).

Therefore the IM test can be interpreted as a test of a null hypothesis moment condition against one for the alternative hypothesis. This overcomes one of the perceived weaknesses of the IM test.

The above results also suggest why IM tests can have poor power. The IM test as commonly used does not vary with possible alternative hypotheses. It is clear from (2.6) or (2.10) that the IM test can be generalized to test (2.9) against alternatives other than $\Sigma_{ij}(X_t,\theta)\cdot g_{ij}(X_t,\theta)\cdot\alpha_{ij}$, thereby increasing the power of the IM test in certain directions.[3]

2.3 A Random Parameter Heterogeneity Interpretation of the IM Test

Tests for random parameter heterogeneity have been given by Chesher (1984) and Cox (1983). Chesher considered the general framework of section 2.1, i.e. heterogeneity in the q parameters θ, and found that a score test against quite general forms of heterogeneity in θ is equivalent to the IM test. Cox instead considered heterogeneity in the underlying parameters η. He focused on ~~considered only~~ the simplest specialization of the type in section 2.2, namely that the underlying parameter η is a scalar, without loss of generality equal to the mean of y. Here we present ~~generalize~~ the approach of Cox for ~~to~~ tests of heterogeneity in h underlying parameters.

We consider the following form of random parameter heterogeneity. The functional form for the density function $f(y_t,\cdot)$ is the same under the null and alternative hypotheses. Under the null hypothesis, the underlying parameter vector of this density is $\eta(X_t,\theta)$, as defined in section 2.2. Under the alternative hypothesis the parameter of this density is random with mean $\eta(X_t,\theta)$ and variance matrix of $O(T^{-1/2})$ that is defined below.

We have the following hypotheses for the density:

$$(2.14) \qquad H_0\text{: } f(y_t,\lambda_t), \quad \lambda_t = \eta(X_t,\theta) \ ,$$

$$(2.15) \qquad H_1\text{: } f(y_t,\lambda_t), \quad \lambda_t \text{ random with mean } \eta(X_t,\theta), \text{ variance matrix } \Gamma_t,$$

where $\Gamma_t = \Gamma(X_t,\theta,\alpha)$ is a positive definite matrix with typical element:

$$(2.16) \qquad \Gamma_{t,ij}(X_t,\theta,\alpha_{ij}) \;=\; T^{-1/2}\cdot g_{ij}(X_t,\theta)'\cdot\alpha_{ij},$$

where $g_{ij}(X_t,\theta)$ is defined in (2.8), and α_{ij} is a vector of unknown parameters.

Under the null hypothesis of no random parameter heterogeneity, $\Gamma_t = 0$. This can be tested by testing whether $\alpha_{ij} = 0$, $i = 1,..,j$, $j = 1,..,h$. In the Appendix, an $O(T^{-1})$ approximation to the density of y_t conditional on $\eta(X_t,\theta)$ and $\Gamma_t(X_t,\theta,\alpha)$ is obtained. The score test for $\alpha_{ij} = 0$, using this approximate density, is a test for a specific form of random parameter heterogeneity, viz. that defined by (2.16).

Proposition 2: The score test based on a local approximation to the density (2.15), with random parameter heterogeneity that has specific functional form for the covariance matrix equal to that given in (2.16), is a test of the moment conditions:

$$(2.17) \qquad \mathbb{E}_0[H_{ij}(y_t,\eta(X_t,\theta))\cdot g_{ij}(X_t,\theta) \mid X_t] = 0 \ . \qquad \blacksquare$$

The derivation of Proposition 2 is given in the Appendix.

The moment restrictions (2.17) equal those tested by the IM test, given in (2.6). It follows that the IM test is a test of a very particular form of random parameter heterogeneity, namely heterogeneity in the h underlying parameters with variance matrix defined in (2.16). This is to be contrasted with the result of Chesher (1984), that the IM test can be interpreted as a score test for quite general forms of heterogeneity in the q parameters θ.

In applications, random parameter heterogeneity is typically posited for the underlying parameters, the approach taken here. The resulting generalizations of the null hypothesis model are mixture models. In such applications, the assumed functional form for the variance matrix of the random parameters is typically much simpler than that in (2.17). For example, the negative binomial model for count data results when the (random) mean parameter for the Poisson is gamma distributed, and in applications the variance of the (random) mean parameter is simply a multiple of its mean or the square of its mean; see Hausman, Hall, and Griliches (1984).

2.4 Implementation of IM Tests

The above discussion has focused on the function $d(y_t,X_t,\theta)$ to be used in forming the sample moment $d_T(\theta)$ in (2.4). In implementing an IM test, θ needs to be replaced by an estimator $\hat{\theta}_T$, consistent under H_0.

Auxiliary regressions to implement IM tests are given in Chesher (1983) and Lancaster (1984). For example, Lancaster (1984) shows that T times the uncentered R^2 from regression of 1 on $d(y_t,X_t,\hat{\theta}_T)$ and $\nabla_\theta L(y_t,X_t,\hat{\theta}_T)$ is chi-square with $q(q+1)/2$ degrees of freedom under H_0. The more general literature on implementation of CM tests may also be used.[4]

As noted by White (1982, p.12) considerable simplification occurs if $E_0[\nabla_\theta d(y_t,X_t,\theta) \mid X_t] = 0$, since then the asymptotic distribution of $d_T(\theta)$ in (2.4) is unaffected by replacing θ by $\hat{\theta}_T$. For example, we can modify Lancaster's procedure and regress 1 on $d(y_t,X_t,\hat{\theta}_T)$.

For the IM test given in section 2.2, we note that such simplifications are possible if:

$$(2.18) \qquad E_0[\ \nabla_\theta(H_{ij}(y_t,X_t,\theta)) \mid X_t] \ = \ 0, \quad i = 1,\ldots,j, \quad j = 1,\ldots,h.$$

Then a chi-squared test statistic is computed as T times uncentered R^2 from the auxiliary OLS regression of 1 on $g_{ij}(X_t,\hat{\theta}_T)\cdot H_{ij}(y_t,X_t,\hat{\theta}_T)$. Following Cameron and Trivedi (1990), given (2.11) an asymptotically equivalent chi-square test statistic can be computed as T times uncentered R^2 from the OLS regression:

$$(2.19) \qquad \Sigma_{ij}(X_t,\hat{\theta}_T)^{-1/2}\cdot H_{ij}(y_t,X_t,\hat{\theta}_T) = \Sigma_{ij}(X_t,\hat{\theta}_T)^{1/2}\cdot g_{ij}(X_t,\hat{\theta}_T)'\cdot\alpha_{ij} + \text{error}.$$

The regression (2.19) actually permits testing the significance of individual components of $g_{ij}(X_t,\theta)$ by 't-tests' on individual components of α_{ij}. Note that it is just the weighted least squares regression given in section 2.2, with θ replaced by $\hat{\theta}_T$. Condition (2.18) holds for a number of examples given below.

3. Examples

3.1 IM Tests for Regression Models based on the Linear Exponential Family

The linear exponential family (LEF) is a one-parameter distribution that generates a wide range of commonly used models. The LEF includes the normal (with known variance), Poisson, binomial (with known number of trials), gamma, exponential, and geometric distributions. For more details, see Gourieroux, Montfort, and Trognon (1984) and McCullagh and Nelder (1989). For CM tests for the LEF, see Wooldridge (1990b).

Using the mean parameterization, the LEF is defined by:

$$(3.1) \qquad f(y,\mu) = \exp\{A(\mu) + B(y) + C(\mu)\cdot y\} ,$$

where the functions A, B and C are such that the density integrates to 1, and it can be shown that:

$$(3.2a) \qquad \mathbb{E}[y] \equiv \mu = -(\nabla_\mu C(\mu))^{-1}\cdot\nabla_\mu A(\mu)$$

$$(3.2b) \qquad \mathbb{E}[(y-\mu)^2] \equiv V(\mu) = (\nabla_\mu C(\mu))^{-1}$$

$$(3.2c) \qquad \mathbb{E}[(y-\mu)^3] = V(\mu)\cdot\nabla_\mu V(\mu)$$

$$(3.2d) \qquad \mathbb{E}[(y-\mu)^4] = V(\mu)\{(\nabla_\mu V(\mu))^2 + V(\mu)\cdot\nabla_\mu^2 V(\mu) + 3V(\mu)\} .$$

In terms of the discussion in section 2.2, η equals μ, here a scalar, and regression models are obtained by letting $\mu = \mu(X_t,\theta)$. For example, for the linear regression model $\mu(X_t,\theta) = X_t'\theta$, and for the Poisson regression model, usually $\mu(X_t,\theta) = \exp(X_t'\theta)$. From the Appendix:

$$(3.3) \qquad H_{11}(y_t,X_t,\theta) = \nabla^2_\mu L(y_t,\mu(X_t,\theta)) + (\nabla_\mu L(y_t,\mu(X_t,\theta)))^2$$
$$= V(\mu_t)^{-2}\{(y_t-\mu_t)^2 - \nabla_\mu V(\mu_t)\cdot(y_t-\mu_t) - V(\mu_t)\} \quad ,$$

$$(3.4) \qquad g_{11}(X_t,\theta) = \mathrm{vech}(\nabla_\theta\mu(X_t,\theta)\cdot\nabla_\theta\mu(X_t,\theta)') \quad ,$$

$$(3.5) \qquad \Sigma_{11}(X_t,\theta) = \mathrm{Var}_0(H_{11}(y_t,X_t,\theta))$$
$$= V(\mu_t)^{-2}\cdot(\nabla^2_\mu V(\mu_t) + 2) \quad .$$

The information matrix can therefore be interpreted as a test of the following null versus alternative hypotheses:

$$(3.6) \qquad H_0\text{: } \mathbb{E}_0[V(\mu_t)^{-2}\{(y_t-\mu_t)^2 - \nabla_\mu V(\mu_t)\cdot(y_t-\mu_t) - V(\mu_t)\} \mid X_t] = 0 \ ,$$

$$(3.7) \qquad H_1\text{: } \mathbb{E}_1[V(\mu_t)^{-2}\{(y_t-\mu_t)^2 - \nabla_\mu V(\mu_t)\cdot(y_t-\mu_t) - V(\mu_t)\} \mid X_t]$$
$$= V(\mu_t)^{-2}\cdot(\nabla^2_\mu V(\mu_t) + 2)\cdot\mathrm{vech}(\nabla_\theta\mu(X_t,\theta)\cdot(\nabla_\theta\mu(X_t,\theta))')'\cdot\alpha_{11} \ .$$

This can also be interpreted as a test, given $\mathbb{E}_0[y_t \mid X_t] = \mathbb{E}_1[y_t \mid X_t] = \mu_t$, of the following null versus alternative hypotheses:

(3.8) $\quad H_0\colon \; E_0[(y_t-\mu_t)^2 \mid X_t] \; = \; V(\mu_t) \quad ,$

(3.9) $\quad H_1\colon \; E_1[(y_t-\mu_t)^2 \mid X_t]$

$$= V(\mu_t) \; + \; (\nabla_\mu^2 V(\mu_t) + 2)\cdot \text{vech}(\nabla_\theta\mu(X_t,\theta)\cdot(\nabla_\theta\mu(X_t,\theta))')'\cdot\alpha_{11}.$$

Thus in regression models based on the LEF, the IM test is a test, conditional on correct specification of the conditional mean, of a particular form of misspecification of the conditional variance function, namely that given in the second line of (3.9).

To connect this with the random parameter heterogeneity interpretation given in section 2.3, note that if y_t has LEF density with mean parameter λ_t that is random with mean μ_t and variance $g_{11}(X_t,\theta)'\cdot\alpha_{11}$, then conditional on μ_t and α_{11}, y_t has mean μ_t and variance $E[V(\lambda_t)] + g_{11}(X_t,\theta)'\cdot\alpha_{11}$.[5] But a locally equivalent alternative is that y_t has mean μ_t and variance $V(\mu_t) + g_{11}(X_t,\theta)'\cdot\alpha_{11}$, i.e. the alternative in (3.9).[6]

Implementation of the IM test in the LEF case is straightforward. It follows from (3.3) that $E_0[\nabla_\mu H_{11}(y,\mu) \mid X] \; = \; 0$, and hence:

$$E_0[\nabla_\theta H_{11}(y_t,X_t,\theta) \mid X_t] \; = \; 0 \quad . \tag{3.10}$$

So (2.18) is satisfied, and we can simply use T times uncentered R^2 or individual t-tests from the OLS regression:

(3.11) $\quad V(\hat\mu_t)^{-1}\cdot(\nabla_\mu^2 V(\hat\mu_t) + 2)^{-1/2}\cdot\{(y_t-\hat\mu_t)^2 - \nabla_\mu V(\hat\mu_t)\cdot(y_t-\hat\mu_t) - V(\hat\mu_t)\}$

$$= V(\hat\mu_t)^{-1}\cdot(\nabla_\mu^2 V(\hat\mu_t) + 2)^{1/2}\cdot\text{vech}(\nabla_\theta\mu(X_t,\hat\theta_T)\cdot(\nabla_\theta\mu(X_t,\hat\theta_T))')'\cdot\alpha_{11} + \text{error},$$

where $\hat\mu_t = \mu(X_t,\hat\theta_T)$, and $\hat\theta_T$ is the MLE for θ_0 under H_0.

For example, for the Poisson regression model, $V(\mu_t) = \mu_t$, so $\nabla_\mu V(\mu_t) =$

1, $\Sigma_t = 2\mu_t^{-1}$, and usually $\mu_t = \exp(X_t'\theta)$, so $\nabla_\theta \mu(X_t'\theta) = \mu_t X_t$ and $g_{11} = \mu_t^2 \cdot \mathrm{vech}(X_t X_t')$. So the IM test, seemingly without an alternative hypothesis, can be interpreted as a test for overdispersion of the form $\mathrm{Var}(y_t \mid X_t) = V(\mu_t) + \mathrm{vech}(X_t X_t')' \cdot \alpha_{11}$, and the test is easily implemented by OLS regression of $(\hat{\mu}_t)^{-1}\{(y_t - \hat{\mu}_t)^2 - y_t\}$ on $(\hat{\mu}_t)^{-1}\mathrm{vech}(X_t X_t')$.

Similar interpretation and implementation is as straightforward for other models in the LEF, which will only vary in the choices of variance function $V(\mu_t)$ and regression function $\mu(X_t, \theta)$.

3.2 IM Tests for Nonlinear Regression Models based on the Normal Distribution

As an example of a regression model based on a two-parameter distribution, we consider the non-linear regression model, y_t is normal with mean $\mu_t = \mu(X_t, \beta)$ and variance σ^2. This example is that of Hall (1987), except that he considered the linear regression model. Using results in Appendix A.4, and letting the subscripts 11, 22, and 12 denote the terms w.r.t. μ, σ^2 and the cross-product, the alternative hypotheses H_1: $\mathbb{E}_1[H_{ij}] = \Sigma_{ij} \cdot g_{ij}' \cdot \alpha_{ij}$ are:

(3.12a) H_1: $\mathbb{E}_1[\{(y_t - \mu_t)^2 - \sigma^2\} \mid X_t] = \mathrm{vech}(\nabla_\beta \mu(X_t, \beta) \cdot (\nabla_\beta \mu(X_t, \beta))')' \cdot \alpha_{11}$

(3.12b) H_1: $\mathbb{E}_1[\{(y_t - \mu_t)^3 - 3\sigma^2 (y_t - \mu_t)\} \mid X_t] = \nabla_\beta \mu(X_t, \beta)' \cdot \alpha_{12}$

(3.12c) H_1: $\mathbb{E}_1[\{(y_t - \mu_t)^4 - 6\sigma^2 (y_t - \mu_t)^2 + 3\sigma^4\} \mid X_t] = \alpha_{22}$,

and the null hypotheses are that each of the functions on the left-hand side have expectation zero. Equivalently, the IM test can be interpreted as successive tests, given $\mathbb{E}_0[y_t \mid X_t] = \mathbb{E}_1[y_t \mid X_t] = \mu_t$, of the following null versus alternative hypotheses:

$$(3.13)\quad H_0:\ E_0[(y_t-\mu_t)^2 \mid X_t] = \sigma^2 ,$$

$$(3.14)\quad H_1:\ E_1[(y_t-\mu_t)^2 \mid X_t] = \sigma^2 + \text{vech}(\nabla_\beta\mu(X_t,\beta)\cdot(\nabla_\beta\mu(X_t,\beta))')'\cdot\alpha_{11} ,$$

$$(3.15)\quad H_0:\ \mathbb{E}_0[(y_t - \mu_t)^3 \mid X_t] = 0 ,$$

$$(3.16)\quad H_1:\ \mathbb{E}_1[(y_t - \mu_t)^3 \mid X_t] = \nabla_\beta\mu(X_t,\beta)'\cdot\alpha_{12} ,$$

$$(3.17)\quad H_0:\ \mathbb{E}_0[(y_t - \mu_t)^4 \mid X_t] = 3\sigma^4 ,$$

$$(3.18)\quad H_1:\ \mathbb{E}_1[(y_t - \mu_t)^4 \mid X_t] = 3\sigma^4 + \alpha_{22} .$$

It is well-known that the IM test in this case is a test of heteroskedasticity, symmetry and non-normal kurtosis. See, for example, White (1982) and Hall (1987). The following is a more precise statement. In the nonlinear regression model with homoskedastic normally distributed error, the IM test decomposes into tests of: (1) a particular form of heteroskedasticity, namely that in (3.14), given correct specification of the conditional mean; (2) a particular form of symmetry, namely that in (3.16), given correct specification of the first two conditional moments; and (3) a particular form of non-normal kurtosis, namely that in (3.18), given correct specification of the first three conditional moments.

Implementation of the IM test in this case is straightforward. We can simply use T times uncentered R^2 or individual t-tests from the OLS regressions:

$$(3.19a)\quad (y_t - \hat{\mu}_t)^2 - \hat{\sigma}^2 = \text{vech}(\nabla_\beta\mu(X_t,\beta)\cdot\nabla_\beta\mu(X_t,\beta)')'\cdot\alpha_{11} + \text{error}$$ FN

$$(3.19b)\quad (y_t - \hat{\mu}_t)^3 - 3\hat{\sigma}^2(y_t - \hat{\mu}_t) = \nabla_\beta\mu(X_t,\beta)'\cdot\alpha_{12} + \text{error}$$

$$(3.19c)\quad (y_t - \hat{\mu}_t)^4 - 6\hat{\sigma}^2(y_t - \hat{\mu}_t)^2 + 3\hat{\sigma}^4 = \alpha_{22} + \text{error} ,$$

FN $E_0[\nabla_{\sigma^2}\{(y_t - \mu_t)^2 - \sigma^2\}] = -1 \neq 0$, but regression (3.19a) is valid provided it includes a constant term.

since the functions H_{ij} in the left-hand side of (3.12a)-(3.12c) have zero derivative w.r.t. β, so that (2.18) holds.

3.3 Discussion

Most applications of the IM test have been to the normal regression model. In this case the null hypothesis is that (second through fourth) moments of residuals are constant, and the alternative hypothesis is that these moments are respectively quadratic, linear or constant functions of the regressors.

The analysis of this paper indicates that generalization of this special case is possible, though is perhaps not immediate. In place of moments of residuals, we have the expectation of the products and cross-products of the sum of the second derivative and outer product of the first derivatives of the density, where these derivatives are taken with respect to the underlying parameters of the density, i.e. $H_{ij}(Y_t,X_t,\theta)$ defined in (2.7). Under the alternative hypothesis, the non-zero expectation equals a linear function of the cross product of the derivatives of the underlying parameters with respect to the model parameters. Unlike the normal regression model case, this linear function of cross products under the alternative is additionally weighted by a variance function, given in general by (2.12), and for the LEF case in (3.9).

In general, the function $H_{ij}(Y_t,X_t,\theta)$ may not be readily interpretable, but it is in the examples given above, which subsume many commonly-used econometric models, and in some additional models. In particular, the LEF results can be immediately extended to some non-LEF regression models, models that conditional on some unknown "nuisance" parameters are LEF. If an IM test is performed conditional on these nuisance parameters, then the above theory is still correct to (3.10). However, the tests cannot in general be implemented using the regression (3.11), since (2.18) is not satisfied with respect to the nuisance parameters.[7] Yet another application of the IM test is that to the truncated normal model, given by Orme (1990). By inspection of equations (4.8) of Orme (1990, p.318), the functions $H_{ij}(Y_t,X_t,\theta)$ are then functions of the first four truncated moments of the data.

For the normal regression model, Hall (1987) noted that the three components for the IM test in the normal regression model are equivalent to score tests for heteroskedasticity (Breusch and Pagan (1979)), and skewness and non-normal kurtosis (Bera and Jarque (1982)). And for a number of members of the LEF family, the IM tests in section 3.2 are equivalent to score tests proposed for departures from the null hypothesis variance-mean relationship, see Cameron (1990).

This coincidence of the IM tests with score tests generalizes, in that the IM test can be interpreted as a score test of random parameter heterogeneity of a particular type, namely that with variance of the underlying parameters given by (2.16). However, in general there is no reason to believe that for a given model the score tests that are equivalent to the IM test will be the score tests typically used to test specification of the given model.

4. Extensions

The preceding analysis has been of the simplest variant of the IM test, the partial IM test which drops the second term in (A.3). This term is identically zero if

$$(3.20) \qquad \nabla^2_\theta \eta_i(X_t,\theta) = 0 , \quad i = 1,\ldots,h.$$

Condition (3.20) holds for linear parameterizations $\eta_i(X_t,\theta) = X_{it}'\theta_i$, including the even simpler $\eta_i = \theta_i$. This is the case for the linear regression model, Hall (1987), and its extensions to truncation, Orme (1990), and autoregressive errors, Bera and Lee (1990). For these examples the theory of section 2 holds with no qualification.

Examples of IM tests where $\eta_i(X_t,\theta)$ is not linear in θ_i, in which case (3.20) does not hold, are seldom studied. The preceding analysis is still appropriate if we view it as the IM test conditional on

$$(3.21) \qquad \mathbb{E}_0[\nabla_{\eta_i} L(y_t,\eta(X_t,\theta)) \mid X_t] = 0 ,$$

for those i for which $\eta_i(X_t'\theta)$ is non-linear in θ_i. For non-linear regression models under normality and for LEF regression models, (3.21) holds if the conditional mean of y_t is correctly specified, as noted at the end of Appendix sections A.3 and A.4. So for these examples the theory of section 2 and discussion in section 3 holds with the qualification that it is assumed that the conditional mean of the dependent variable is correctly specified.

Alternatively, we can consider a "full" IM test which includes the second term in (A.3). Then from the Appendix, Proposition 1 is unchanged for the cross terms $i \neq j$. For the case $i = j$, Proposition 1 is modified by adding an additional term $\mathrm{vech}(\nabla^2_{\theta_i\theta_i}\eta_i(X_t,\theta_i))\cdot\nabla_{\eta_i} L(y_t,\eta(X_t,\theta))$ in (2.6). An interpretation along the lines of the Corollary is in general no longer possible.

Simplification is possible if the only non-linearity in $\eta_i(X_t,\theta_i)$ is that it is a non-linear transformation of a linear function in θ_i, viz. $\eta_i(X_t,\theta_i) = \eta_i(X_{it}'\theta_i)$. Then, from the Appendix, the only change to section 2.2, i.e. both Proposition 1 and its Corollary, is that $H_{ij}(y_t,X_t,\theta)$ is replaced by:

(3.22) $H_{ij}^*(y_t,X_t,\theta) = H_{ij}(y_t,X_t,\theta) + (\nabla\eta_i)^{-2}\cdot\nabla^2\eta_i\cdot\nabla_{\eta_i}L(y_t,\eta(X_t,\theta))$,

where $\nabla^k\eta_i$ denotes $\nabla^k_{X_{it}'\theta_i}\eta_i(X_{it}'\theta_i)$, $k = 1,2$.

This type of nonlinearity in $\eta_i(X_t,\theta_i)$ is that typically used for LEF regression models, such as discrete choice models and Poisson models.[8] In that case $i = 1$, $\eta_i(X_t,\theta) = \mu(X_t'\theta)$, and from Appendix A.3 $\nabla_\mu L(y_t,\mu_t) = V(\mu_t)^{-1}\cdot(y_t - \mu_t)$. The full IM test will therefore be a test on a particular linear combination of the first two conditional moments of y_t.

The random parameter heterogeneity interpretation of the IM test of section 2.3 can also be modified to take account of the second term in A.3. As shown in the Appendix, the null hypothesis (2.14) is unchanged, but the alternative hypothesis is generalized - the parameter is random with conditional mean not necessarily equal to $\eta(X_t,\theta)$. Specifically, (2.15) becomes λ_t is random with mean of λ_{it} equal to $\eta_i(X_{it},\theta_i)$ + $T^{-1/2}\cdot\text{vech}(\nabla^2_{\theta_i\theta_i}\eta_i(X_t,\theta_i))'\alpha_{ii}$, and mean-square error of λ_t equal to Γ_t defined in (2.16). The full IM test is equivalent to a score test against this form of random parameter heterogeneity.

Interpretation of the IM test is simplest if (3.20) or (3.21) holds. Previous in-depth studies of IM tests for specific examples have implicitly restricted analysis to only cases where (3.20) holds, by considering the linear regression model. And while a term analogous to the second term in (A.3) appears in Chesher (1984), assumptions made permit neglect of this term.[9] So the issue of whether or not to assume (3.21) has not usually arisen.

The difference between imposing or not imposing (3.21) is well illustrated by IM testing for the discrete choice model.[10] In this model y_t is Bernoulli distributed, a member of the LEF (binomial with one trial) with mean μ_t and variance function $V(\mu_t) = \mu_t(1 - \mu_t)$. Then $H_{11}(y_t,X_t,\theta)$ in

(3.3) equals $(\mu_t(1-\mu_t))^{-2}\cdot\{(y_t-\mu_t)^2 - (1 - 2\mu_t)\cdot(y_t-\mu_t) - \mu_t(1-\mu_t)\}$, which equals zero when y_t takes either of its possible values 0 and 1. Thus imposing (3.21) leaves nothing to test.[11]

Instead we should additionally use the second term in (A.3), equal to $\nabla^2_\theta\mu(X_t,\theta)\cdot(\mu_t(1-\mu_t))^{-1}\cdot(y_t - \mu_t)$. Then the full IM test for the discrete choice models is a test of the hypothesis:

$$(3.23) \qquad H_0\colon\ \mathbb{E}_0[\mathrm{vech}(\nabla^2_\theta\mu(X_t,\theta))\cdot(\mu_t(1-\mu_t))^{-1}\cdot(y_t - \mu_t) \mid X_t] = 0 .$$

Using $\mathrm{Var}(y_t) = \mu_t(1-\mu_t)$, this can be interpreted as the optimal RBCM test of:

$$(3.24) \qquad H_0\colon\ \mathbb{E}_0[y_t \mid X_t] = 0 ,$$

against the alternative hypothesis:

$$(3.25) \qquad H_0\colon\ \mathbb{E}_0[y_t \mid X_t] = \mathrm{vech}(\nabla^2_\theta\mu(X_t,\theta))'\alpha .$$

To summarize, for the discrete choice model the partial IM test is not applicable, while the full IM test is a test for a specific form of misspecification of the conditional mean.

4. Conclusion

Parametric regression models are typically based on densities depending on few, say h, parameters, where these few parameters are in turn modeled to depend on explanatory variables and many, say q, unknown parameters. Two types of IM test, sometimes equivalent, can be considered in this case: a partial IM test and a full IM test.

The partial IM test can be interpreted as a test of whether just $h(h+1)/2$ functions of the dependent variable have expectation zero, against the alternative that they have expectation equal to the variance of the function of the dependent variable multiplied by a linear combination of elements of the outer product of the derivative of the underlying h parameters with respect to the q unknown parameters.

Thus the partial IM test can in general be interpreted as a test of a null hypothesis moment condition against an alternative hypothesis moment condition. Furthermore, the claim that the IM test is in general a test for random parameter heterogeneity of unspecified form in the q unknown parameters can be tightened, to a claim that it is a test for local random parameter heterogeneity with a specific functional form for the covariance matrix of the underlying parameters.

In considering specific examples, previous studies have taken the following approach: obtain the IM test, find that it coincides with a particular well-known score test, and hence conclude that the IM test is a test against whatever form of misspecification the equivalent score test was testing against. The interpretation in this paper is less circuitous, as it instead directly obtains the underlying null and alternative hypothesis moment conditions.

Because previous studies have focused on the linear regression model, they have implicitly restricted analysis to the case where the partial IM test coincides with the full IM test. The more general case where the full IM test differs from the partial IM test is also analyzed. In this case the full IM test again coincides with a score test for a very specific, albeit different, functional form for local random parameter heterogeneity.

Footnotes

[1] If some of the moment conditions (2.3) are already imposed in estimation, they are omitted from the IM test.

[2] The IM test considered here is the original White (1982) IM test, called the Second-Order IM test by White (1990). Two other IM tests are the Cross-Information Matrix test, White (1990), and, for dynamic models, the Dynamic (First-Order) IM test, White (1987).

[3] IM tests based only on a subset of the elements of g_{ij} in (2.8) have been considered by some authors. Wooldridge (1990b) proposes CM tests that are variants of IM tests, using regressors other than g_{ij}.

[4] References include Newey (1985), Tauchen (1985), Pagan and Vella (1989), White (1990), Wooldridge (1990a). Much of this literature focuses on ways to implement CM tests by means of auxiliary (or artificial) regressions.

[5] We use $\mathrm{Var}(y) = E[\mathrm{Var}(y|\lambda)] + \mathrm{Var}(E[y|\lambda]) = E[V(\lambda)] + \mathrm{Var}(\lambda)$.

[6] Since for local alternatives $\alpha_{11} = T^{-1/2}\gamma_{11}$, $E[V(\lambda_t)] = V(\mu_t) + o_p(T^{-1})$.

[7] Even with nuisance parameters, regression (3.10), can sometimes be run. This is the case for the normal, and for the negative binomial if the variance is an unknown scalar multiple of the mean. In other cases, more general implementation methods need to be used, with the results of Wooldridge (1990a) particularly relevant. For the LEF with nuisance parameters, Wooldridge (1990b) proposed CM specification tests based on $\{(y_t-\mu_t)^2 - V(\mu_t)\}$, rather than $H_{11}(y_t,X_t,\theta)$ in (3.3), so that the simplest computational procedure, running regression (3.11), is not used for the LEF models.

[8] The common discrete choice models are the probit and logit models, where the conditional mean is respectively $\Phi(X_t'\beta)$ and $L(X_t'\beta)$, and Φ and L are respectively the standard normal and logistic c.d.f.'s. For the Poisson, the conditional mean is usually parameterized as $\exp(X_t'\beta)$.

[9] The second term in the second line of equation (2) of Chesher (1984) disappears if the matrix R is constant across observations.

[10] This example arose from conversation with Dick Jefferis.

[11] This result should not be surprising, as for data taking only two values, 0 or 1, the only possible distribution is the Bernoulli, and the only possible distributional misspecification is in the Bernoulli parameter. (3.21) in this example implies that the Bernoulli parameter is correctly specified, leaving no other misspecification to test.

APPENDIX

A.1 Derivation of Proposition 1

Suppose the density of the vector dependent variable y_t conditional on explanatory variables X_t is of the form $f(y_t,X_t,\theta) = f(y_t,\eta(X_t,\theta))$, where $\eta(X_t,\theta)$ is a hx1 vector function of the qx1 parameter vector θ such that θ is identified in the sense that $\eta(X_t,\theta_1) = \eta(X_t,\theta_2)$ iff $\theta_1 = \theta_2$. Then $L(y_t,\eta(X_t,\theta)) = \log f(y_t,\eta(X_t,\theta))$ has first derivative:

$$\begin{aligned}\nabla_\theta L(y_t,\eta(X_t,\theta)) &= \nabla_\theta(\eta(X_t,\theta)')\cdot\nabla_\eta L(y_t,\eta(X_t,\theta)) \\ &= \sum_{i=1}^{h} \nabla_\theta\eta_i(X_t,\theta)\cdot\nabla_{\eta_i} L(y_t,\eta(X_t,\theta)) \quad ,\end{aligned}$$

where η_i is the i-th component of η. Differentiating again, (2.1) becomes

$$\text{(A.1)}\qquad \begin{aligned}D(y_t,X_t,\theta) &= \nabla_\theta(\eta(X_t,\theta)')\cdot H(y_t,X_t,\theta)\cdot(\nabla_\theta(\eta(X_t,\theta)'))' \\ &\quad + \sum_{i=1}^{h} \nabla_\theta^2\eta_i(X_t,\theta)\cdot\nabla_{\eta_i} L(y_t,\eta(X_t,\theta)) \quad ,\end{aligned}$$

where

$$\text{(A.2)}\qquad H(y_t,X_t,\theta) = \nabla_\eta^2 L(y_t,\eta(X_t,\theta)) + \nabla_\eta L(y_t,\eta(X_t,\theta))\cdot\nabla_\eta L(y_t,\eta(X_t,\theta))' \quad ,$$

where $\nabla_\eta^2 \equiv \partial^2/\partial\eta\partial\eta'$. Then the IM test is based on the qxq matrix of moment conditions:

$$\text{(A.3)}\qquad \begin{aligned}&E_0[\nabla_\theta(\eta(X_t,\theta)')\cdot H(y_t,X_t,\theta)\cdot(\nabla_\theta(\eta(X_t,\theta)'))' \\ &\quad + \sum_{i=1}^{h} \nabla_\theta^2\eta_i(X_t,\theta)\cdot\nabla_{\eta_i} L(y_t,\eta(X_t,\theta)) \mid X_t] = 0 \quad .\end{aligned}$$

Different variants of the IM test arise according to whether or not the second term in (A.3) is zero. Initially we consider the simplest case where the second term is dropped, either by doing analysis conditional on the assumption $E_0[\nabla_\eta L(y_t,\eta(X_t,\theta)) \mid X_t] = 0$, or because $\nabla^2_\theta \eta_i(X_t,\theta) = 0$, $i = 1,\ldots,h$. This latter condition holds in the most studied case, the IM test for the classical linear regression model under normality.

Vectorizing, the IM test is based on the q^2x1 vector of moment conditions, not necessarily unique:

$$\text{(A.4)} \qquad E_0[(\nabla_\theta \eta(X_t,\theta)' \otimes \nabla_\theta \eta(X_t,\theta)')\cdot \text{vec}(H(y_t,X_t,\theta)) \mid X_t] = 0 \quad .$$

Now impose the additional structure on $\eta(X_t,\theta)$, that it equals $(\eta_1(X_t,\theta_1)' \cdots \eta_h(X_t,\theta_h)')'$, where θ_i, i=1,...,h, are non-overlapping q_ix1 components of the qx1 vector θ, $q = q_1 + \cdots + q_h$. Then the IM test based on the unique elements of (A.4) is a test of the moment conditions:

$$\text{(A.5)} \quad E_0[\text{vec}(\nabla_{\theta_i} \eta_i(X_t,\theta)\cdot(\nabla_{\theta_j} \eta_j(X_t,\theta))')\cdot H_{ij}(y_t,X_t,\theta) \mid X_t] = 0, \qquad i \neq j,$$

$$E_0[\text{vech}(\nabla_{\theta_i} \eta_i(X_t,\theta)\cdot(\nabla_{\theta_i} \eta_i(X_t,\theta))')\cdot H_{ii}(y_t,X_t,\theta) \mid X_t] = 0, \qquad i = j,$$

$$i = 1,\ldots,j, \quad j = 1,\ldots,h,$$

where $H_{ij}(y_t,X_t,\theta)$ is the (i,j)-th entry of $H(y_t,X_t,\theta)$.

A.2 Derivation of Proposition 2

For simplicity we initially suppress the subscript t. The null hypothesis model is that conditional on η, y has density $f(y,\eta)$. The alternative hypothesis model is that conditional on λ, y has density $f(y,\lambda)$, but λ is itself a random variable with density $p(y,\eta,\Gamma)$, where conditional on η and Γ, λ has mean η and variance-covariance matrix Γ. Following Cox (1983), we consider modest amounts of parameter heterogeneity, by assuming that Γ is $O(T^{-1/2})$.

We obtain the following approximation of the density of y, conditional on η and Γ:

$$
\begin{aligned}
\text{(A.6)} \qquad g(y,\eta,\Gamma) &= \int f(y,\lambda)p(\lambda,\eta,\Gamma)\, d\lambda \\
&= \int \{f(y,\eta) + \nabla_\eta f(y,\eta)' \cdot (\lambda-\eta) \\
&\qquad + \tfrac{1}{2}(\lambda-\eta)' \nabla_\eta^2 f(y,\eta) \cdot (\lambda-\eta) + O(T^{-1})\} \cdot p(\lambda,\eta,\Gamma)\, d\lambda \\
&= f(y,\eta) + 0 + \frac{1}{2}\,\mathrm{tr}(\nabla_\eta^2 f(y,\eta) \cdot \Gamma) + O(T^{-1}) \\
&= f(y,\eta) + \frac{1}{2}\,\mathrm{tr}(f(y,\eta) \cdot H(y,\eta) \cdot \Gamma) + O(T^{-1}) \\
&= f(y,\eta) \cdot [1 + \frac{1}{2}\,\mathrm{tr}(H(y,\eta) \cdot \Gamma) + O(T^{-1})] \\
&= f(y,\eta) \cdot \exp[\frac{1}{2}\,\mathrm{tr}(H(y,\eta) \cdot \Gamma)] + O(T^{-1})
\end{aligned}
$$

where

$$
\text{(A.7)} \qquad H(y,\eta) = \nabla_\eta^2 L(y,\eta) + \nabla_\eta L(y,\eta) \cdot \nabla_\eta L(y,\eta)' ,
$$

with $L(y,\eta) = \log(f(y,\eta))$. In (A.6), the second line follows from a second-order Taylor series expansion of $f(y,\lambda)$ about $\lambda = \eta$; the third-line follows from the assumed mean and variance of λ given η and Γ; the fourth line uses the result that for any density under suitable regularity conditions $\nabla_\eta^2 f(y,\eta) = f(y,\eta) \cdot H(y,\eta)$, since $\nabla_\eta^2 L(y,\eta) = f(y,\eta)^{-1} \cdot \nabla_\eta^2 f(y,\eta) - \nabla_\eta L(y,\eta) \cdot \nabla_\eta L(y,\eta)'$; and the final line uses the approximation $\exp(x) = 1 + x$ for small x.

Now suppose the random parameter vector λ_t in the density $f(y_t,\lambda_t)$ has mean vector $\eta(X_t,\theta)$, where $\eta(X_t,\theta)$ is as defined in section 2.2, and variance matrix Γ_t with typical element $\Gamma_{t,ij} = T^{-1/2} \cdot g_{ij}(X_t,\theta) \cdot \alpha_{ij}$, where $g_{ij}(X_t,\theta)$ is defined in (2.8), and α_{ij} is a vector of unknown parameters.

Under the null hypothesis of no random parameter heterogeneity, $\alpha_{ij} = 0$, $i = 1,..,j$, $j = 1,..,q$. From (A.6):

$$\text{(A.8)} \qquad \log(g(y_t,\eta_t,\Gamma_t)) = \log(f(y_t,\eta_t)) + \frac{1}{2}\operatorname{tr}(H(y_t,\eta_t)\cdot\Gamma_t) + O(T^{-1}) \ .$$

Differentiating w.r.t. α_{ij}, which appears only in $\Gamma_{t,ij}$, the score test for $\alpha_{ij} = 0$ is based on the moment conditions:

$$\text{(A.9)} \qquad E_0[H_{ij}(y_t,\eta(X_t,\theta))\cdot g_{ij}(X_t,\theta) \mid X_t] = 0 \ .$$

But (A.9) equal the moment restrictions tested by the IM test, and given in (2.6).

A.3 IM Tests for LEF Models

From (3.1), $L(y,\mu) = A(\mu) + B(y) + C(\mu)\cdot y$, so that:

$$\begin{aligned}
\text{(A.10)} \qquad \nabla_\mu L(y,\mu) &= \nabla_\mu A(\mu) + \nabla_\mu C(\mu)\cdot y \\
&= \nabla_\mu C(\mu)\cdot(y - (\nabla_\mu C(\mu))^{-1}\cdot\nabla_\mu A(\mu)) \\
&= V(\mu)^{-1}\cdot(y - \mu) \quad ,
\end{aligned}$$

where the last line follows from (3.2a) and (3.2b). Hence:

$$\begin{aligned}
\text{(A.11)} \qquad H_{11}(y,\mu) &\equiv \nabla^2_\mu L(y,\mu) + (\nabla_\mu L(y,\mu))^2 \\
&= \{-V(\mu)^{-2}\nabla_\mu V(\mu)\cdot(y-\mu) - V(\mu)^{-1}\} + \{V(\mu)^{-2}\cdot(y-\mu)^2\} \\
&= V(\mu)^{-2}\cdot\{(y-\mu)^2 - \nabla_\mu V(\mu)\cdot(y-\mu) - V(\mu)\} \quad .
\end{aligned}$$

Squaring, taking expectations, and using equations (3.2), (A.12) $\quad \mathrm{Var}(H_{11}(y,\mu))$

$$= V(\mu)^{-4}[V(\mu)\{(\nabla_\mu V(\mu))^2 + V(\mu)\cdot\nabla_\mu^2 V(\mu) + 3V(\mu)\} + (\nabla_\mu V(\mu))^2\cdot(V(\mu))$$
$$+ (V(\mu))^2 - 2\,\nabla_\mu V(\mu)\cdot V(\mu)\cdot\nabla_\mu V(\mu) - 2\,V(\mu)\cdot V(\mu)]$$
$$= V(\mu)^{-4}[V(\mu)^2\cdot\nabla_\mu^2 V(\mu) + 2\,(V(\mu))^2] \;.$$
$$= V(\mu)^{-2}\cdot[\nabla_\mu^2 V(\mu) + 2] \;.$$

The second term in (A.3) that was dropped equals $\nabla_\theta^2\mu(X,\theta)\cdot V(\mu)^{-1}\cdot(y - \mu)$. This has expectation zero if the conditional mean of y is correctly specified.

A.4 IM Tests for Normal Models

$$L(y,\mu,\sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - \mu)^2 \quad \text{for the } N(\mu,\sigma^2)$$

density, so that

$$\text{(A.13a)} \quad \nabla_\mu L(y,\mu,\sigma^2) = \frac{1}{\sigma^2}(y - \mu) \;,$$
$$\text{(A.13b)} \quad \nabla_{\sigma^2} L(y,\mu,\sigma^2) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y - \mu)^2 = \frac{1}{2\sigma^4}\{(y - \mu)^2 - \sigma^2\} \;,$$
$$\text{(A.13c)} \quad \nabla_\mu^2 L(y,\mu,\sigma^2) = -\frac{1}{\sigma^2} \;,$$
$$\text{(A.13d)} \quad \nabla_{\sigma^2}^2 L(y,\mu,\sigma^2) = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(y - \mu)^2 = \frac{1}{4\sigma^8}\{2\sigma^4 - 4\sigma^2(y - \mu)^2\} \;,$$
$$\text{(A.13e)} \quad \nabla_{\mu,\sigma^2}^2 L(y,\mu,\sigma^2) = -\frac{1}{2\sigma^4}(y - \mu) \;,$$

and hence H_{ij}, g_{ij}, and Σ_{ij} are respectively:

$$\text{(A.14a)} \quad H_{11}(y,\mu,\sigma^2) = \nabla_\mu^2 L(y,\mu,\sigma^2) + (\nabla_\mu L(y,\mu,\sigma^2))^2$$
$$= \frac{1}{\sigma^4}\cdot\{(y - \mu)^2 - \sigma^2\} \;,$$
$$\text{(A.14b)} \quad H_{12}(y,\mu,\sigma^2) = \nabla_{\mu,\sigma^2}^2 L(y,\mu,\sigma^2) + \nabla_{\sigma^2} L(y,\mu,\sigma^2)\cdot\nabla_\mu L(y,\mu,\sigma^2)$$
$$= \frac{1}{2\sigma^6}\cdot\{(y - \mu)^3 - 3\sigma^2(y - \mu)\} \;,$$

(A.14c) $$H_{22}(y,\mu,\sigma^2) = \nabla^2_{\sigma^2} L(y,\mu,\sigma^2) + (\nabla_{\sigma^2} L(y,\mu,\sigma^2))^2$$
$$= \frac{1}{4\sigma^8}\cdot\{(y-\mu)^4 - 6\sigma^2(y-\mu)^2 + 3\sigma^4\} \quad ,$$

(A.15a) $$g_{11}(X,\beta,\sigma^2) = \mathrm{vec}^*(\nabla_\beta\mu(X,\beta)\cdot\nabla_\beta\mu(X,\beta)')$$

(A.15b) $$g_{12}(X,\beta,\sigma^2) = \nabla_\beta\mu(X,\beta)$$

(A.15c) $$g_{22}(X,\beta,\sigma^2) = 1 \quad ,$$

(A.16a) $$\Sigma_{11}(X,\beta,\sigma^2) = \mathrm{Var}_0(H_{11}(y,\mu,\sigma^2)) = \frac{2}{\sigma^4}$$

(A.16b) $$\Sigma_{12}(X,\beta,\sigma^2) = \mathrm{Var}_0(H_{12}(y,\mu,\sigma^2)) = \frac{3}{2\sigma^6}$$

(A.16c) $$\Sigma_{22}(X,\beta,\sigma^2) = \mathrm{Var}_0(H_{22}(y,\mu,\sigma^2)) = \frac{3}{2\sigma^8} .$$

It follows that the alternative hypotheses $H_1 : \mathbb{E}_1[H_{ij}] = \Sigma_{ij}\cdot g_{ij}\cdot\alpha_{ij}$ are:

(A.17a) $$\mathbb{E}[\frac{1}{2\sigma^4}\{(y-\mu)^2 - \sigma^2\} \mid X] = \frac{2}{\sigma^4}\cdot\mathrm{vec}^*(\nabla_\beta\mu(X,\beta)\cdot\nabla_\beta\mu(X,\beta)')'\cdot\alpha_{11}$$

(A.17b) $$\mathbb{E}[\frac{1}{2\sigma^6}\{(y-\mu)^3 - 3\sigma^2(y-\mu)\} \mid X] = \frac{3}{2\sigma^6}\cdot\nabla_\beta\mu(X,\beta)\cdot\alpha_{12}$$

(A.17c) $$\mathbb{E}[\frac{1}{4\sigma^8}\{(y-\mu)^4 - 6\sigma^2(y-\mu)^2 + 3\sigma^4\} \mid X] = \frac{3}{2\sigma^8}\cdot\alpha_{22} \quad .$$

The equations in the text follow directly, upon canceling the common multiples in σ and constants.

The second term in (A.3) is zero for i = 2, and for i = 1 is equal to $\frac{1}{\sigma^2}\cdot(y-\mu)'\cdot\nabla^2_\theta\mu(X,\theta)$. This has expectation zero if the conditional mean of y is correctly specified, and in the linear regression model is zero since $\nabla^2_\theta\mu(X,\theta) = \nabla^2_\theta(X'\theta) = 0$.

The above results are readily extended to the case where the variance is specified to be heteroskedastic with functional form $\sigma^2 = v(X,\gamma)$. Then the only changes will be that (A.15b) and (A15.c) become $g_{12}(X,\beta,\gamma)$ = $\text{vec}^*(\nabla_\beta \mu(X,\beta) \cdot \nabla_\gamma v(X,\beta)')$ and $g_{22}(X,\beta,\gamma) = \text{vec}^*(\nabla_\gamma v(X,\gamma) \cdot \nabla_\gamma v(X,\gamma)')$, with corresponding changes to the right-hand sides of (A.17b) and (A.17c).

A.5 Modification of Proposition 1 for the General Case

We consider the case where the second term in (A.3) does not disappear. For the density defined in (2.5), i.e. $\eta(X_t,\theta) = (\eta_1(X_t,\theta_1)' \cdot\cdot\ \eta_h(X_t,\theta_h)')'$, where θ_i are non-overlapping components of θ, the (i,j)-th subcomponent of (A.3) yields the $q_i x q_j$ matrix of moment conditions:

$$\text{(A.18)} \qquad E_0[\nabla_{\theta_i}\eta_i(X_t,\theta_i)\cdot(\nabla_{\theta_j}\eta(X_t,\theta_j))'\cdot H_{ij}(y_t,X_t,\theta) + \nabla^2_{\theta_i\theta_j}\eta_i(X_t,\theta_i)\cdot\nabla_{\eta_i}L(y_t,\eta(X_t,\theta)) \mid X_t] = 0 \quad .$$

For cross terms, the second term in (A.16) will always disappear, since $\nabla^2_{\theta_i\theta_j}\eta_i(X_t,\theta_i) = 0$ for $i \neq j$, and we obtain the same result as in (A.5). The only concern is the case i = j. Then vectorizing yields:

$$\text{(A.19)} \qquad E_0[\text{vech}(\nabla_{\theta_i}\eta_i(X_t,\theta_i)\cdot(\nabla_{\theta_j}(\eta(X_t,\theta_j))')\cdot H_{ij}(y_t,X_t,\theta) + \text{vech}(\nabla^2_{\theta_i\theta_j}\eta_i(X_t,\theta_i))\cdot\nabla_{\eta_i}L(y_t,\eta(X_t,\theta)) \mid X_t] = 0 \quad .$$

This differs from (A.5) in the second term, which is an orthogonality condition involving the first derivative of the log-likelihood function with respect to the underlying parameter.

Simplification is possible if $\eta_i(X_t,\theta_i) = \eta_i(X_{it}'\theta_i)$. Then $\nabla_{\theta_i}\eta_i(X_t'\theta_i) = \nabla\eta_i \cdot X_t$ and $\nabla^2_{\theta_i}\eta_i(X_t,\theta_i) = \nabla^2\eta_i \cdot X_t X_t'$, where $\nabla^k\eta_i$ denotes $\nabla^k_{X_{it}'\theta_i}\eta_i(X_{it}'\theta_i)$, k = 1,2. (A.19) simplifies to:

$$(A.20)\qquad E_0[\mathrm{vech}(\nabla\eta_i\cdot X_t)\cdot(\nabla\eta_i\cdot X_t)')\cdot H_{ij}(y_t,X_t,\theta) + \mathrm{vech}(\nabla^2\eta_i\cdot X_tX_t'\cdot\nabla_{\eta_i}L(y_t,\eta(X_t,\theta)) \mid X_t] = 0 \quad .$$

Equivalently:

$$(A.21)\qquad E_0[\mathrm{vech}(\nabla\eta_i\cdot X_t)\cdot(\nabla\eta_i\cdot X_t)'\cdot\{H_{ij}(y_t,X_t,\theta) + (\nabla\eta_i)^{-2}\cdot\nabla^2\eta_i\cdot\nabla_{\eta_i}L(y_t,\eta(X_t,\theta)) \mid X_t] = 0 \quad ,$$

which leads to (3.22).

A.6 Modification of Proposition 2 for the General Case

Proceed as in section A.2, with the modification that the mean of λ equals $(\eta + \gamma)$, rather than η, where γ is $O(T^{-1/2})$ and Γ is now the mean-square error of λ. Then the second term in the third equality in (A.6), previously zero, becomes $\nabla_\eta f(y,\eta)'\cdot\gamma$, and equation (A.6) becomes

$$(A.22)\qquad g(y,\eta,\gamma,\Gamma) = f(y,\eta)\cdot\exp[\nabla_\eta L(y,\eta)'\cdot\gamma + \frac{1}{2}\,\mathrm{tr}(H(y,\eta)\cdot\Gamma)] + O(T^{-1}) \ .$$

Parameterize Γ_t as before, and let the mean of the i-th component of λ_t equal $\eta_i(X_{it},\theta_i) + T^{-1/2}\cdot\mathrm{vech}(\nabla^2_{\theta_i\theta_i}\eta_i(X_t,\theta_i))'\alpha_{ii}$, which implies that $\gamma_{it} = T^{-1/2}\cdot\mathrm{vech}(\nabla^2_{\theta_i\theta_i}\eta_i(X_t,\theta_i))'\alpha_{ii}$. Then (to a multiple $T^{-1/2}$)

$$(A.23)\qquad \nabla_{\alpha_{ii}}\log g(y,\eta,\gamma,\Gamma) = H_{ij}(y_t,\eta(X_t,\theta))\cdot g_{ij}(X_t,\theta) + \mathrm{vech}(\nabla^2_{\theta_i\theta_i}\eta_i(X_t,\theta_i))\cdot\nabla_{\eta_i}L(y_t,\eta(X_t,\theta)),$$

so that the score test coincides with the IM test given in (A.19).

References

Bera, A.K. and C.M. Jarque, 1982, "Model Specification Tests: A Simultaneous Approach", *Journal of Econometrics,* 20, 59–82.

Bera, A.K. and S. Lee, 1990, "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis", Working Paper 90-26, Department of Economics, U.C.-San Diego.

Breusch, T.S. and A.R. Pagan, 1979, "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica,* 47, 1287–1294.

Cameron, A.C., 1990, "Regression-Based Tests of Variance-Mean Relationships", forthcoming working paper, Department of Economics, U.C.-Davis.

Cameron, A.C. and P.K. Trivedi, 1990, "Conditional Moment Tests with Explicit Alternatives", Working Paper #366, Department of Economics, U.C.-Davis.

Chesher, A., 1983, "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation", *Economics Letters,* 13, 45–48.

Chesher, A., 1984, "Testing for Neglected Heterogeneity", *Econometrica,* 52, 865–872.

Cox, D.R., 1983, "Some Remarks on Overdispersion", *Biometrika*, 70, 269–274.

Gourieroux, C., A. Monfort and A.Trognon, 1984, Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 52, 3, 681–700.

Hall, A., 1987, "The Information Matrix Test in the Linear Model", *Review of Economic Studies,* 54, 257–265.

Hausman, J., B.H. Hall and Z. Griliches, 1984, "Econometric Models for Count Data with an Application to the Patents - R&D Relationship, *Econometrica*, 52, 909–938.

Lancaster, T., 1984, "The Covariance Matrix of the Information Matrix Test", *Econometrica,* 52, 1051–1054.

McCullagh, P. and J. A. Nelder, 1989, *Generalized Linear Models*, 2nd ed., *London*: Chapman and Hall.

Newey, W.K., 1985, "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047–1070.

Orme, C., 1990, "The Small-Sample Properties of the Information-Matrix Test", *Journal of Econometrics*, 46, 309–331.

Pagan, A.R. and F. Vella, 1989, "Diagnostic Tests for Models Based on Individual Data", *Journal of Applied Econometrics,* Supplement, S4, S29–S60.

Tauchen, G. 1985, "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, 30, 415–443.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1–25.

White, H., 1984, "Comment on Tests of Specification in Econometrics", *Econometrics Reviews*, 3, 261–267.

White, H., 1987, "Specification Testing in Dynamic Models", in T. Bewley, ed., *Advances in Econometrics, Fifth World Congress*, 1, 1–58, Cambridge University Press, New York.

White, H., 1990, *Estimation, Inference and Specification Analysis,* Cambridge University Press, forthcoming.

Wooldridge, J.M., 1990a, "A Unified Approach to Robust, Regression-Based Specification Tests", *Econometric Theory*, 6, 17–43.

Wooldridge, J.M., 1990b, "Specification testing and Quasi-likelihood Estimation", *Journal of Econometrics*, forthcoming.