THE INFORMATION MATRIX TEST:
ITS COMPONENTS AND IMPLIED ALTERNATIVE HYPOTHESES

By

A. Colin Cameron
Department of Economics
University of California
Davis, CA, 95616, U.S.A.

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

June 1992

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. By comparison to the most studied example of an IM test, the linear regression model with homoskedastic normally distributed error, the IM test will in general test more fundamental forms of misspecification. As an illustration, IM tests for regression models based on the linear exponential family are presented and interpreted.

Some Key Words: information matrix tests; conditional moment specification tests; score tests; linear exponential family.

JEL Classification: C12, C52

Acknowledgements: This paper has benefited from presentation at the Universities of Rochester, Illinois, and California - Davis, Indiana University, the 1992 Winter Meetings of the Econometric Society and Camp Econometrics IV. Discussion with Anil Bera, Rob Feenstra, Dick Jefferis and Adrian Pagan has been especially helpful.

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). Furthermore, in empirical applications the IM test can be applied to models for which there is as yet no standard battery of specification tests. 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 some cases the IM test for a specific model coincides with an existing test, in practice usually a score test. Then the IM test is interpreted as a test of the same form of misspecification as that for the existing test. The best known example is the linear regression model with homoskedastic normally distributed error, discussed in White (1982, p.12) and analyzed in detail by Hall (1987). Then the IM test is a test of specific forms of departure from homoskedasticity, symmetry and normal kurtosis, which for inference on the regression parameters are second-order effects compared to misspecification of the conditional mean.

In other cases, however, the IM test is a test of more fundamental forms of misspecification. Less well known is that for regression models under normality that are nonlinear, the IM test is additionally a test of misspecification of the conditional mean of the dependent variable.

In this paper we provide a general interpretation of the IM test, one that additionally explains why an IM test in different settings can be testing such radically different forms of misspecification. The key step is to recognize that 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 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 general theory is presented in section 2. In section 3 we present and interpret IM tests for regression models based on the linear exponential family. These regression models are based on only one underlying parameter. The results are discussed further in section 4, and some concluding remarks are made in section 5.

2. General Theory

2.1 The Information Matrix Test

We model 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 $q\text{x}1$ 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.[1]

Define

$$(2.1) \qquad s_\theta(y_t, X_t, \theta) = \nabla_\theta \log f(y_t, X_t, \theta)$$

to be the score for the t-th observation, and

$$(2.2) \qquad D(y_t, X_t, \theta) = \nabla_{\theta'} s_\theta(y_t, X_t, \theta) + s_\theta(y_t, X_t, \theta) \cdot s_\theta(y_t, X_t, \theta)' ,$$

[1] Extension to time series is straightforward by conditioning on $\{y_{t-1}, y_{t-2}, \ldots, X_t, X_{t-1}, X_{t-2}, \ldots\}$ rather than X_t alone.

where ∇_θ and $\nabla_{\theta'}$ denote derivatives with respect to θ and θ'. For densities satisfying the regularity conditions, 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 $f(y_t,X_t,\theta)$. The information matrix (IM) test of White (1982) is a test of the $q(q+1)/2$ unique moment conditions[2]:

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

where vech(•) is the "vector half" operator which stacks the lower triangular part of a symmetric matrix into a column vector.

We note that the IM Test is a special case of the conditional moment (CM) tests of Newey (1985) and Tauchen (1985). In this more general framework, $\text{vech}(D(y_t,X_t,\theta))$ in (2.3) may be replaced by any function with expectation zero under the assumed model, not just that function defined by (2.2). The CM test framework is used below.

The IM test is implemented by testing the departure from zero of the corresponding sample moment, $T^{-1}\sum_{t=1}^{T} \text{vech}(D(y_t,X_t,\hat{\theta}_T))$, where $\hat{\theta}_T$ is an estimator consistent for θ under the true model. In this paper we are concerned with interpretation of the moment condition (2.3), rather than methods of implementation.

---

[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). Sometimes (2.3) includes moment restrictions already imposed by estimation. If so, these would be omitted.

2.2 The IM Test in Terms of Underlying Parameters

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 parameter 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 $q = \dim(\beta) + 1$. In other commonly used univariate regression models h is also of low dimension, typically 1 or 2.

The assumed density is therefore of the form $f(y_t,\eta(X_t,\theta)) = f(y_t,\eta_1(X_t,\theta_1),\ldots,\eta_h(X_t,\theta_h))$, or more simply:

$$f(y_t,\eta_t) = f(y_t,\eta_{1t},\ldots,\eta_{ht}) , \tag{2.4}$$

where $\eta_t = \eta(X_t,\theta)$ does not depend on y_t.

To obtain the IM test given the density (2.4), we first define

$$s_\eta(y_t,\eta_t) = \nabla_\eta \log f(y_t,\eta(X_t,\theta)) \tag{2.5}$$

to be the score with respect to the underlying parameters η, and

$$H(y_t,\eta_t) = \nabla_{\eta'} s_\eta(y_t,\eta_t) + s_\eta(y_t,\eta_t)s_\eta(y_t,\eta_t)' \tag{2.6}$$

to be the matrix which satisfies the information matrix equality in terms of the underlying parameters η. Then by application of the chain rule of differentiation, (2.1) becomes

$$(2.7) \qquad s_\theta(y_t, X_t, \theta)) = \nabla_\theta \eta_t' \cdot s_\eta(y_t, \eta_t) = \sum_{i=1}^{h} \nabla_\theta \eta_{it} \cdot s_{\eta_i}(y_t, \eta_t) \quad .$$

Differentiating again, (2.2) becomes after some rearrangement:

$$(2.8) \qquad D(y_t, X_t, \theta) = \sum_{i=1}^{h} \nabla_\theta^2 \eta_{it} \cdot s_{\eta_i}(y_t, \eta_t) + \nabla_\theta \eta_t' \cdot H(y_t, \eta_t) \cdot (\nabla_\theta \eta_t')'$$

where ∇_θ^2 denotes the second derivative.

The IM test is a test of $\mathbb{E}_0[\mathrm{vec}(D(y_t, X_t, \theta)) \mid X_t] = 0$. From (2.8), the dependent variable y_t appears in two ways. First it appears as the score vector with respect to the underlying parameters η. Second it appears in the term $H(y_t, \eta_t)$ which has expectation zero by the information matrix equality for the model in the underlying parameters η. We directly obtain the following proposition.

Proposition 1

The IM test is a conditional moment test that simultaneously tests:

$$(2.9) \qquad \mathbb{E}_0[\sum_{i=1}^{h} \mathrm{vec}(\nabla_\theta^2 \eta_{it}) \cdot s_{\eta_i}(y_t, \eta_t) \mid X_t] = 0$$

$$(2.10) \qquad \mathbb{E}_0[(\nabla_\theta \eta_t')' \otimes (\nabla_\theta \eta_t')' \cdot \mathrm{vec}(H(y_t, \eta_t)) \mid X_t] = 0 \; . \qquad \blacksquare$$

By the law of iterated expectation, the IM test is accordingly a simultaneous test of the following forms of misspecification: (1) that the score vector for the model in the underlying parameters does not have zero mean, i.e. $\mathbb{E}[s_\eta(y_t, \eta_t) \mid X_t] \neq 0$; and (2) that the score vector for the model in the underlying parameters does not satisfy the information matrix equality: i.e. $\mathbb{E}[H(y_t,\eta_t) \mid X_t] \neq 0$.

An interesting question is in which directions (2.9) and (2.10) are testing the two forms of misspecification. At first glance the directions appear to be exactly $(\nabla_\theta \eta_t')'\otimes(\nabla_\theta \eta_t')'$, the nonlinear analog of squares and products of regressors, and $\sum_{i=1}^{h} \nabla_\theta^2 \eta_{it}$, respectively. But this needs qualification as given in the next section.

2.3 An Underlying Alternative for the IM Test

We impose the additional structure that for $\eta(X_t,\theta) = (\eta_1(X_t,\theta_1)' \cdots \eta_h(X_t,\theta_h)')'$, the θ_i are non-overlapping components of θ, $q = q_1 + \cdots + q_h$.[3] Then the (i,j)-th subcomponent of (2.8) associated with the underlying parameters η_i and η_j is of dimension $q_i x q_j$ and is given by

$$(2.11) \quad D^{ij}(y_t,X_t,\theta) = \nabla^2_{\theta_i\theta_j}\eta_{it}\cdot s_{\eta_i}(y_t,\eta_t) + \nabla_{\theta_i}\eta_{it}\cdot H_{ij}(y_t,\eta_t)\cdot(\nabla_{\theta_j}\eta_{jt})',$$

where $H_{ij}(y_t,\eta_t) = \nabla_{\eta_i} s_{\eta_j}(y_t,\eta_t) + s_{\eta_i}(y_t,\eta_t)s_{\eta_j}(y_t,\eta_t)$ is the (i,j)-th entry in $H(y_t,\eta_t)$.

The first term in (2.11) will always disappear for cross-terms, since $\nabla^2_{\theta_i\theta_j}\eta_i(X_t,\theta_i) = 0$ for $i \neq j$. However, it will not disappear when $i = j$, unless $\eta_i(X_t,\theta_i) = X_t'\theta_i$, in which case $\nabla^2_{\theta_i}\eta_i(X_t,\theta_i) = 0$. The second term will always be present, unless $H_{ij}(y_t,\eta_t) = 0$. Therefore upon vectorization

---

[3] The assumption of non-overlapping components of θ is satisfied in usual applications of the IM test, but does exclude, for example, multivariate models with cross-equation parameter restrictions. The approach taken here can be appropriately modified on a case by case basis.

the (i,j)-th subcomponents corresponding to Proposition 1 are:

(2.12) $$\mathbb{E}_0[s_{\eta_i}(y_t,\eta_t)\cdot \text{vech}(\nabla^2_{\theta_i}\eta_{it}) \mid X_t] = 0 , \quad i = j ,$$

(2.13) $$\mathbb{E}_0[H_{ii}(y_t,\eta_t)\cdot \text{vech}(\nabla_{\theta_i}\eta_{it}\cdot(\nabla_{\theta_i}\eta_{jt})') \mid X_t] = 0 , \quad i = j,$$

$$\mathbb{E}_0[H_{ij}(y_t,\eta_t)\cdot \text{vec}(\nabla_{\theta_i}\eta_{it}\cdot(\nabla_{\theta_j}\eta_{jt})') \mid X_t] = 0 , \quad i \neq j.$$

To further interpret (2.12) and (2.13) we use the following construct.

The moment condition (2.13) is of the form $\mathbb{E}_0[H(y_t,\eta_t)\cdot g(\eta_t) \mid X_t] = 0$, where $H(\cdot)$ is a scalar, $g(\cdot)$ is a vector and η_t is not a function of y_t. Then tests of this moment condition are "WLS regression based" conditional moment tests of H_0: $\mathbb{E}_0[H(y_t,\eta_t) \mid X_t] = 0$ against H_1: $\mathbb{E}_1[H(y_t,\eta_t) \mid X_t] = \text{Var}(H(y_t,\eta_t))\cdot g(\eta_t)'\alpha$. The rationale is that H_1 suggests regression of $H(y_t,\eta_t)$ on $\text{Var}(H(y_t,\eta_t))\cdot g(\eta_t)$. Since $H(y_t,\eta_t)$ may be heteroskedastic the most powerful test of H_0: $\alpha = 0$ against local alternatives $\alpha = T^{-1/2}\gamma$ will use the weighted least squares (WLS) estimator which divides terms by $\text{Var}(H(y_t,\eta_t))^{1/2}$. Thus

$$\hat{\alpha}_{WLS} = \{\sum_{t=1}^{T} g(\eta_t)\cdot\text{Var}(H(y_t,\eta_t))^{-1}\cdot g(\eta_t)'\}^{-1}\cdot\sum_{t=1}^{T} g(\eta_t,\theta)\cdot H(y_t,\eta_t) .$$

Tests based on $\hat{\alpha}_{WLS}$ 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 $\mathbb{E}_0[H(y_t,\eta_t)\cdot g(\eta_t) \mid X_t] = 0$.

This approach is discussed in more detail in Cameron and Trivedi (1990a). The key part of this interpretation is that tests of $\mathbb{E}_0[H(y_t,\eta_t)\cdot g(\eta_t) \mid X_t] = 0$ are tests of $\mathbb{E}_1[H(y_t,\eta_t) \mid X_t] = 0$ against $\mathbb{E}_1[H(y_t,\eta_t) \mid X_t] = \text{Var}(H(y_t,\eta_t))\cdot g(\eta_t)'\alpha$ rather than the more obvious $g(\eta_t)'\alpha$. Unless $H(y_t,\eta_t)$

is homoskedastic these alternatives differ. A similar interpretation is used for the moment condition (2.12).

Proposition 2

Let θ_i in $\eta_{it} = \eta_i(X_t,\theta_i)$ be non-overlapping components of $\theta = (\theta_1',..,\theta_h')'$. Then the first component of the IM test is zero for $i \neq j$, and for $i = j$ is a WLS regression based CM test of:

$$(2.14) \qquad H_0\colon\ \mathbb{E}_0[s_{\eta_i}(y_t,\eta_t) \mid X_t] \ = \ 0\ ,$$

against

$$(2.15) \qquad H_1\colon\ \mathbb{E}_1[s_{\eta_i}(y_t,\eta_t) \mid X_t] \ = \ \mathrm{Var}(s_{\eta_i}(y_t,\eta_t))\cdot \mathrm{vech}(\nabla^2_{\theta_i}\eta_{it})'\alpha_{1ii}\ .$$

The second component of the IM test is a WLS regression based CM test of:

$$(2.16) \qquad H_0\colon\ \mathbb{E}_0[H_{ij}(y_t,\eta_t) \mid X_t] \ = \ 0\ ,$$

against

$$(2.17) \qquad H_1\colon\ \mathbb{E}_1[H_{ij}(y_t,\eta_t) \mid X_t]$$

$$= \ \mathrm{Var}(H_{ii}(y_t,\eta_t))\cdot \mathrm{vech}(\nabla_{\theta_i}\eta_{it}\cdot(\nabla_{\theta_i}\eta_{jt})')'\alpha_{2ii},\quad i = j,$$

$$= \ \mathrm{Var}(H_{ij}(y_t,\eta_t))\cdot \mathrm{vech}(\nabla_{\theta_i}\eta_{it}\cdot(\nabla_{\theta_j}\eta_{jt})')'\alpha_{2ij},\quad i \neq j\ . \ \blacksquare$$

Therefore the IM test can be interpreted as a test of a null hypothesis moment condition against one for the alternative hypothesis, with the alternative hypothesis model parameterized by α in addition to θ.[4] This overcomes one of the perceived weaknesses of the IM test.

---

[4] Of course (2.15) and (2.17) are just one representation of the alternatives, as there will be many locally equivalent alternatives.

The above results also suggest why IM tests can have poor power. First, the implicit null hypotheses are that $s_{\eta_i}(y_t,\eta_t)$ and $H_{ij}(y_t,\eta_t)$ have expectation zero. These are just two of many functions of y_t and the underlying parameters η_t that might be chosen as the basis for a test.

Second, the directions of departure under the implied alternatives given in Proposition 2 are very specific. For example, if $\eta_{it} = X_{it}'\theta_i$, i.e. a linear specification is chosen for the underlying parameter, the first part of the IM test drops out completely, and for the second part the direction is $\mathrm{Var}(H_{ij}(y_t,\eta_t))\cdot\mathrm{vech}(X_{it}\cdot X_{jt}')'\alpha_2$. The power of the IM test might be improved by testing in other directions. Furthermore, we might test in fewer than the q(q+1)/2 directions of the IM test. Thus CM tests of $\mathbb{E}_0[s_{\eta_i}(y_t,\eta_t)\cdot g_i(X_t) \mid X_t] = 0$, and $\mathbb{E}_0[H_{ij}(y_t,\eta_t)\cdot g_{ij}(X_t) \mid X_t] = 0$ might be implemented, where $g_i(X_t)$ and $g_{ij}(X_t)$ are of lower dimension than the corresponding terms in (2.12) and (2.13). Obvious candidates are powers of η_{it} and η_{jt}.

Third, the IM test does not test the two conditions (2.9) and (2.10) separately, sequentially or jointly. Instead the sum of the two conditions is tested, as given in (2.8). Since it is possible that one of the two terms of (2.8) may be negative while the other is positive, the IM test may not detect misspecification when separate tests of the two conditions would. Separate CM tests of the two conditions might instead be implemented.

3. IM Tests for Regression Models based on the Linear Exponential Family

3.1 IM Test for LEF Models

A wide range of commonly used models are actually based on densities with just one underlying parameter, notably members of the linear exponential family (LEF). The LEF includes the normal (with known variance), binomial (with known number of trials), Poisson, 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 (1991).

We begin by deriving a general expression for the IM test in LEF models. 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:

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

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

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

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

and the score

$$(3.3) \qquad s_\mu(y,\mu) = \nabla_\mu \log(y,\mu) = V(\mu)^{-1}\cdot(y - \mu) .$$

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 it is customary to specify $\mu(X_t,\theta) = \exp(X_t'\theta)$.

The essential property of the LEF that yields the results below is that the score vector in terms of the underlying parameters equals the "residual" $(y - \mu)$ divided by the variance as given in (3.3). Then

$$(3.4) \qquad \begin{aligned} H(y,\mu) &= \nabla_\mu s_\mu(y,\mu) + (s_\mu(y,\mu))^2 \\ &= V(\mu)^{-2}\{(y-\mu)^2 - \nabla_\mu V(\mu)\cdot(y-\mu) - V(\mu)\} , \end{aligned}$$

with variance

$$\text{(3.5)} \qquad \mathrm{Var}(H(y,\mu)) = V(\mu)^{-2}\cdot(\nabla_\mu^2 V(\mu_t) + 2) \ ,$$

obtained by squaring (3.4), taking expectations, using equations (3.2), and simplifying. The IM test is therefore a test of the moment condition:

$$\text{(3.6)} \qquad \mathbb{E}_0[V(\mu_t)^{-1}\cdot(y_t - \mu_t)\cdot \mathrm{vech}(\nabla_\theta^2 \mu_t) + V(\mu_t)^{-2}\{(y_t-\mu_t)^2 - \nabla_\mu V(\mu_t)\cdot(y_t-\mu_t) - V(\mu_t)\} \cdot \mathrm{vech}(\nabla_\theta \mu_t \cdot (\nabla_\theta \mu_t)') \mid X_t] = 0 \ .$$

For applications to any LEF model one can directly use (3.6). Here we are concerned with interpreting the resulting IM test. By proposition 2, the first component of the IM test is a CM test of:

$$\text{(3.7)} \qquad H_0\text{:}\ \mathbb{E}_0[V(\mu_t)^{-1}(y_t-\mu_t) \mid X_t] = 0 \ ,$$

against

$$\text{(3.8)} \qquad H_1\text{:}\ \mathbb{E}_1[V(\mu_t)^{-1}(y_t-\mu_t) \mid X_t] = V(\mu_t)^{-1}\cdot \mathrm{vech}(\nabla_\theta^2 \mu_t)'\alpha_1 \ ,$$

and the second component of the IM test is a test of

$$\text{(3.9)} \qquad H_0\text{:}\ \mathbb{E}_0[V(\mu_t)^{-2}\{(y_t-\mu_t)^2 - \nabla_\mu V(\mu_t)\cdot(y_t-\mu_t) - V(\mu_t)\} \mid X_t] = 0,$$

against

$$\text{(3.10)} \qquad H_1\text{:}\ \mathbb{E}_1[V(\mu_t)^{-2}\{(y_t-\mu_t)^2 - \nabla_\mu V(\mu_t)\cdot(y_t-\mu_t) - V(\mu_t)\} \mid X_t] = V(\mu_t)^{-2}\cdot(\nabla_\mu^2 V(\mu_t) + 2)\cdot \mathrm{vech}(\nabla_\theta \mu_t \cdot (\nabla_\theta \mu_t)')'\alpha_2 \ .$$

Upon obvious simplification, the first component can alternatively be interpreted as a test of:

$$(3.11)\qquad H_0:\ \mathbb{E}_0[y_t \mid X_t] = \mu_t\ ,$$

against

$$(3.12)\qquad H_1:\ \mathbb{E}_1[y_t \mid X_t] = \mu_t + \mathrm{vech}(\nabla^2_\theta \mu_t)'\alpha_1,$$

while the second component can be interpreted as a test, given (3.11) is not rejected, i.e. given $\mathbb{E}[y_t \mid X_t] = \mu_t$, of the following null versus alternative hypotheses:

$$(3.13)\qquad H_0:\ \mathbb{E}_0[(y_t-\mu_t)^2 \mid X_t] = V(\mu_t)\ ,$$

$$(3.14)\qquad H_1:\ \mathbb{E}_1[(y_t-\mu_t)^2 \mid X_t]$$
$$= V(\mu_t) + (\nabla^2_\mu V(\mu_t) + 2)\cdot\mathrm{vech}(\nabla_\theta\mu_t\cdot(\nabla_\theta\mu_t)')'\cdot\alpha_2.$$

Thus in regression models based on the LEF, the IM test is in general a test of (1) a particular misspecification of the conditional mean, and (2) conditional on correct specification of the conditional mean, a particular form of misspecification of the conditional variance function, namely that given in the second line of (3.14). Rather than sequentially performing these tests, the IM test is a test of the sum of the two.

3.2 Specific LEF Examples

The normal distribution with σ^2 known is a member of the LEF. Then $V(\mu) = \sigma^2$, and simplification occurs because $\nabla_\mu V(\mu) = 0$. In particular, (3.14) simplifies to $\mathbb{E}_1[(y_t-\mu_t)^2 \mid X_t] = \sigma^2 + 2\cdot\mathrm{vech}(\nabla_\theta\mu_t\cdot(\nabla_\theta\mu_t)')'\cdot\alpha_2$, which is locally equivalent to $\mathbb{E}_1[(y_t-\mu_t)^2 \mid X_t] = h(\gamma_0 + Z_t'\gamma_1)$ where $Z_t = \mathrm{vech}(\nabla_\theta\mu_t\cdot(\nabla_\theta\mu_t)')$. This is the alternative for usual tests for heteroskedasticity, with Z_t here chosen to be the products and cross products of $\nabla_\theta \mu(X_t, \theta)$. However, unless $\nabla^2_\theta \mu_t = 0$, i.e. $\mu_t = X_t'\theta$, the IM test is additionally testing the specification of the conditional mean, as in (3.12).

Nonlinear normal models are not all that uncommon. An example is the time series model with AR(1) error: $y_t = X_t'\beta + u_t$ where $u_t = \rho u_{t-1} + \varepsilon_t$ and ε_t is i.i.d. $N(0, \sigma^2)$. Then $\mu_t = \mathbb{E}_0[y_t \mid y_{t-1}, \ldots, X_t, \ldots] = \rho y_{t-1} + X_t'\beta - X_{t-1}'\beta\rho$ is nonlinear in β and ρ. The portion of the IM test corresponding to μ will test the conditional mean as well as the conditional variance, as found by Bera and Lee (1992) who analyzed the IM test for this model.

The normal model with σ^2 unknown is a member of the quadratic exponential family, rather than the LEF. This model is considered in an earlier version of this paper, Cameron and Trivedi (1990b). Then $(\eta_{1t}, \eta_{2t}) = (\mu_t, \sigma^2_t)$, and the afore-mentioned test for heteroskedasticity corresponds to the (1,1) subcomponent of the IM test. The (1,2) subcomponent yields a test of symmetry and the (2,2) subcomponent a test of non-normal kurtosis. These correspond to the results of Hall (1987) for the linear model. However, in the same way that the (1,1) subcomponent will additionally test the specification of the conditional mean if $\mu(X_t, \beta)$ is nonlinear, the (2,2) subcomponent will additionally test the specification of the conditional variance if σ^2_t is parameterized as $\sigma^2(X_t, \gamma)$ nonlinear in γ.

A second well-known example of the LEF is the binary choice model.[5] In this model y_t is Bernoulli distributed, a member of the LEF (binomial with one trial) with mean μ_t and variance function $V(\mu_t) = \mu_t(1 - \mu_t)$. The logit model is the special case $\mu_t = \exp(X_t'\beta)/(1 + \exp(X_t'\beta))$ and for the probit model $\mu_t = \Phi(X_t'\beta)$ where Φ is the standard normal c.d.f. Then $H(y, \mu)$

[5] This example arose from conversation with Dick Jefferis. The IM test for the logit model is given in Newey (1985).

in (3.4) equals $(\mu(1-\mu))^{-2}\cdot\{(y-\mu)^2 - (1-2\mu)\cdot(y-\mu) - \mu(1-\mu)\}$, which equals zero when y_t takes either of its possible values 0 and 1. So the only hypothesis tested is that of correct specification of the conditional mean, $\mathbb{E}_0[y_t \mid X_t] = \mu_t$ against $\mathbb{E}_1[y_t \mid X_t] = \mu_t + \text{vech}(\nabla^2_\theta\mu_t)'\alpha_1$.

A third example of the LEF is the Poisson regression model. Then y_t has mean μ_t and variance $V(\mu_t) = \mu_t$. The IM test is a simultaneous test of correct specification of the conditional mean, with $\mathbb{E}_1[y_t \mid X_t] = \mu_t + \text{vech}(\nabla^2_\theta\mu_t)'\alpha_1$, and correct specification of the conditional variance, with $\text{Var}_1[y_t \mid X_t] = \mu_t + 2\cdot\text{vech}(\nabla_\theta\mu_t\cdot(\nabla_\theta\mu_t)')'\cdot\alpha_2$.

4. Discussion

Most applications of the IM test have been to the linear regression model with homoskedastic normal error. 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 not immediate. The IM test is actually a simultaneous test of two forms of misspecification.

The first form of misspecification is one that is not tested in, and not suggested by, the linear regression example. In nonlinear regression models under normality, and more generally LEF models, this is a test that the conditional mean is correctly specified, against the alternative that in addition terms in the second derivatives of the conditional mean should be included. More generally still, from Proposition 2 the test is one of correct specification of $s_{\eta_i}(y_t,\eta_t)$ against the alternative given in (2.15).

The second form of misspecification can be viewed as a generalization of that found by Hall (1987) for the linear regression model under normality. 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 (2.6). 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 (2.17), and for the LEF case in (3.10). In many models the function $H_{ij}(Y_t, X_t, \theta)$ may not be readily interpretable, though it is in the examples given above, which subsume many commonly-used econometric models.

Interpretation of the IM test is clearly simplest if only the second form of misspecification is tested, which is the case if $\nabla^2_\theta \eta_i(X_t, \theta) = 0$. Previous in-depth studies have implicitly restricted analysis to such cases. This interpretation is still possible if we define IM tests to be conditional on $\mathbb{E}[\nabla_{\eta_i} s(y_t, \eta(X_t, \theta)) \mid X_t] = 0$. But the examples given show that this may be the most important moment condition to test, e.g. for binary choice models imposing this condition would leave nothing to test. And the applied researcher will perform the entire test.

Sub-components of the IM test sometimes coincide with existing score tests. For the normal linear regression model, where only the second type of misspecification is tested, Hall (1987) noted that the IM test is equivalent to score tests for heteroskedasticity (Breusch and Pagan (1979)), skewness and non-normal kurtosis (Bera and Jarque (1982)). And for a number of members of the LEF family, the second type of misspecification tested by the IM test of section 3 is equivalent to a score test of departures from the null hypothesis variance-mean relationship, see Cameron (1991). 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.

5. Summary and 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.

The IM test for such models is testing two types of misspecification: (1) that the score vector for the model in the underlying parameters does not have zero mean, and (2) that the score vector for the model in the underlying parameters does not satisfy the information matrix equality. Furthermore, these two misspecifications are tested in very particular directions, determined by the first and second derivatives of the underlying parameters with respect to the q parameters. A weakness of the IM test is that it considers the sum of these two types of misspecification, even though the two may be offsetting.

Because previous studies of the IM test have focused on the linear regression model, they have implicitly restricted analysis to examples where only the second form of misspecification is tested.

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, 1992, "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis", *Review of Economic Studies*, forthcoming.

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

Cameron, A.C., 1991, "Regression-Based Tests of Heteroscedasticity in Models where the Variance depends on the Mean Relationships", working paper #379, Department of Economics, U.C.-Davis.

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

Cameron, A.C. and P.K. Trivedi, 1990b, "The Information Matrix Test and its Implied Alternative Hypotheses", Working Paper #372, 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.

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.

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.

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., 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., 1991, "Specification Testing and Quasi-likelihood Estimation", *Journal of Econometrics*, 48, 29-55.