AN ALTERNATIVE TYPE OF HETEROGENEITY TESTED BY THE INFORMATION MATRIX TEST

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.

August 11, 1992

ABSTRACT

Chesher (1984) proposed a score test for parameter heterogeneity in model parameters. This score test coincided with the information matrix (IM) test of White (1984), permitting interpretation of the IM test as a general test for random parameter heterogeneity. In this note we propose a score test for an alternative type of heterogeneity, one used often in nonlinear regression models. The IM test coincides with a score test for a very particular specification of this type of heterogeneity. This permits interpretation of the IM test as a score test in very specific directions, which will in general differ from those specified in standard heterogeneity tests.

Some Key Words: information matrix tests; score tests; random parameter heterogeneity; mixture models.

JEL Classification: C12, C52

Acknowledgements: This work has benefited from presentation at the Universities of Rochester, Illinois, and California - Davis, Indiana University and the 1992 Winter Meetings of the Econometric Society.

1. Introduction

The information matrix (IM) test proposed by White (1982) is a model specification test that differs from the classical statistical tests in that no alternative hypothesis model is specified. This failure to specify the direction in which it is testing is perceived to be a weakness of the IM test.

In an influential paper, Chesher (1984) showed that the IM test can be regarded as a score test for random parameter heterogeneity. The particular form of random parameter heterogeneity under the alternative hypothesis is that the parameter vector θ is a random variable, with correctly specified mean and constant variance. Under the null hypothesis this variance is zero. This type of heterogeneity is used in the familiar random coefficient model, Hildreth and Houck (1968), for the linear regression model.

A different type of heterogeneity, however, is often specified in other settings. For example, consider testing for overdispersion in Poisson regression models, Lee (1986) and Cameron and Trivedi (1986). Under the null y_t is exponential with parameter λ_t specified to equal $\exp(X_t'\theta)$. Under the alternative λ_t, rather than θ, is specified to be a random variable, with mean $\exp(X_t'\theta)$ and variance $\sigma^2(\exp(X_t'\theta))^l$, for specified l. More generally, this type of heterogeneity is often used to obtain more general models, called mixture models. Cox (1983) derived a score test for this type of heterogeneity in a wide class of models.

In this note we show, by extension of Cox's results, that the IM test coincides with a score test for a very particular specification of this type of heterogeneity, one that in general will differ from that specified in standard heterogeneity tests.

2. General Theory

2.1 General Framework

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 $q\times 1$ parameter vector, that satisfies the regularity conditions of White (1982). This note focuses on cross-section data, $\{(y_t, X_t),\ t = 1,\ldots,T\}$, independent across t.[1]

While the density $f(y_t, X_t, \theta)$ depends on q parameters, it is typically based on a density depending on underlying parameters of dimension much less than q. These *underlying parameters* are in turn modeled to depend on explanatory variables and the q parameters θ. Specifically, when y_t is actually i.i.d., there is a large menu of density functions of the form $f(y_t, \eta)$, where η is a $h\times 1$ vector. For 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 $q_i\times 1$, and $\theta = (\theta_1', \ldots, \theta_h')'$ is the $q\times 1$ vector of parameters to be estimated.[2]

The assumed density is therefore of the form:

$$(2.1) \qquad f(y_t, X_t, \theta) = f(y_t, \eta_1(X_t, \theta_1), \ldots, \eta_h(X_t, \theta_h)) \ .$$

---

[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.

---

[2] For example, in the classical linear regression model under normality, $y_t \sim N(\eta_{1t}, \eta_{2t})$, where $\eta_{1t} = X_t'\beta$ and $\eta_{2t} = \sigma^2$, so $h = 2$ and $q = 1 + \dim(\beta)$.

2.2 Information Matrix Test

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

$$\mathbb{E}[\operatorname{vech}(\nabla^2_{\theta\theta} L(y_t, X_t, \theta) + \nabla_\theta L(y_t, X_t, \theta)\cdot(\nabla_\theta L(y_t, X_t, \theta))') \mid X_t] = 0 ,$$

where $\operatorname{vech}(\cdot)$ is the "vector half" operator which stacks the lower triangular part of a symmetric matrix into a column vector, and $L(y_t, X_t, \theta)$ is the log-density. The IM test is a statistical test based on the departure from zero of the corresponding sample moments, evaluated at the MLE $\hat{\theta}_T$.

For the density defined in (2.1), the (i,j)-th component of the IM test, i.e. corresponding to differentiation w.r.t. to the vectors θ_i and θ_j, is a test of the $q_i q_j$, or $q_i(q_i+1)/2$ if i=j, moment conditions

$$\begin{aligned} (2.2) \qquad & \mathbb{E}[\operatorname{vec}^*(\nabla^2_{\theta_i\theta_j}\eta_{it}\cdot\nabla_{\eta_i} L(y_t, \eta_t) \; + \\ & \operatorname{vec}^*(\nabla_{\theta_i}\eta_{it}\cdot\{\nabla^2_{\eta_i\eta_j} L(y_t, \eta_t) + \nabla_{\eta_i} L(y_t, \eta_t)\cdot\nabla_{\eta_j} L(y_t, \eta_t)\}\cdot(\nabla_{\theta_j}\eta_{jt})') \\ & \mid X_t] = 0, \qquad i = 1,\dots,h, \quad j = 1,\dots,h, \end{aligned}$$

where $\operatorname{vec}^*(\cdot)$ denotes $\operatorname{vech}(\cdot)$ for $i = j$ and $\operatorname{vec}(\cdot)$ for $i \neq j$. For derivation of (2.2) using the chain rule of differentiation, see the Appendix. Note that in the first term, $\nabla^2_{\theta_i\theta_j}\eta_{it} = 0$, for $i \neq j$.

2.3 A Score Test for Heterogeneity in Underlying Parameters

We consider score tests for local random parameter heterogeneity in the h underlying parameters $\eta_1,\dots,\eta_h$, rather than the q parameters $\boldsymbol{\theta}$. The heterogeneity is of the following type. 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 equals $\eta(X_t,\theta)$. Under the alternative hypothesis the parameter of this density is random with mean $\eta(X_t,\theta) + \gamma_t$, and mean-square error matrix Γ_t. Thus the alternative is that both the mean and the variance of the distribution of the random parameters are misspecified.

An approximation to the alternative hypothesis density under local heterogeneity is given in the following lemma.

Lemma: Let y_t conditional on λ_t have density $f(y_t,\lambda_t)$, where λ_t is a random variable with mean $\eta_t + \gamma_t$ and mean square error Γ_t, where γ_t and Γ_t are $O(T^{-1/2})$. A local approximation to the density of y_t conditional on η_t, γ_t and Γ_t is:

$$(2.3)\quad f^*(y_t,\eta_t,\gamma_t,\Gamma_t) = f(y_t,\eta_t)\cdot\exp[\nabla_\eta L(y_t,\eta_t)'\gamma_t + \frac{1}{2}\,\mathrm{tr}(\nabla^2_\eta\, L(y,\eta) + \nabla_\eta L(y,\eta)\cdot\nabla_\eta L(y,\eta)')\cdot\Gamma_t)] + O(T^{-1})$$

A proof is given in the Appendix. It is an extension of the results of Cox (1983) who considered similar heterogeneity, except that he assumed the mean to be correctly specified, i.e. $\gamma_t = 0$.

To perform a score test of the null hypothesis of no heterogeneity, i.e. $\gamma_t = 0$ and $\Gamma_t = 0$, based on (2.3) we need to parameterize γ_t and Γ_t. For the following particular parameterization the score test coincides with the IM test.

Proposition 1: Let y_t conditional on λ_t have density $f(y_t,\lambda_t)$, where λ_t is a random variable with mean $(\eta_t + \gamma_t)$, where $\eta_{it} = \eta_i(X_t,\theta_i)$ and

$$(2.4)\qquad \gamma_{it} = T^{-1/2}\mathrm{vech}(\nabla^2_{\theta_i\theta_i}\eta_{it}(X_t,\theta_i))'\alpha_{ii}$$

and mean square error Γ_t, where

$$(2.5) \qquad \Gamma_{ijt} = T^{-1/2}\mathrm{vech}(\nabla_{\theta_i}\eta_{it}\cdot(\nabla_{\theta_j}\eta_{jt})')'\alpha_{ij} .$$

Then the score test for no heterogeneity, i.e. H_0: $\alpha_{ij} = 0$, coincides with the information matrix test. ■

A proof is given in the Appendix. Clearly the IM test is a test for a very specific form of heterogeneity. Furthermore, in general it is not only a test of the randomness of λ_t, but also of correct specification of the mean of λ_t, a fundamental form of misspecification.

In some special cases this latter form of misspecification may not be tested by the IM test. It is not tested when the underlying parameters η_{it} are linear in θ, since then $\nabla^2_{\theta_i\theta_i}\eta_{it}(X_t,\theta_i) = 0$ and hence $\gamma_{it} = 0$ in (2.4). This is the case for the linear regression model, but not for nonlinear models.

3. Example

As an example of a nonlinear model we consider the Poisson model for count data. The Poisson density is $f(y,\lambda_1) = \exp(-\lambda_1)\lambda_1{}^y/y!$, and it is customary to let $\lambda_1 = \exp(X'\theta_1)$. The standard test for overdispersion is a score test of the Poisson against the Katz system, given by Lee (1986) and Cameron and Trivedi (1986). The latter show that this test is equivalent to a Cox (1983) score test for H_0: $\lambda_{1t} = \eta_{1t}$, where $\eta_{1t} = \exp(X_t'\theta_1)$, against the alternative that λ_t is random with mean η_{1t} and variance $T^{-1/2}\delta\cdot\eta_{1t}^{l}$, for specified l. By contrast, applying proposition 1 shows that again the IM test is a score test against the alternative that λ_t is random with mean $\eta_{1t} + T^{-1/2}\mathrm{vech}(\eta_{1t}X_tX_t')'\alpha_{11}$ and mean square error $T^{-1/2}\mathrm{vech}(\eta_{1t}^2X_tX_t')'\alpha_{11}$.

In this example the IM test is clearly testing a different specification of heterogeneity than the corresponding standard heterogeneity test. Nonetheless it is possible to make a direct connection with the IM test, as noted by Lee (1986). This is an artifact, however, of the particular parameterization chosen for η_{1t}.

To see this, note that in the special case $\eta_{1t} = \exp(\theta_{11} + X_{2t}'\theta_{12})$, $\nabla^2_{\theta_{11}\theta_{11}}\eta_{1t} = \nabla_{\theta_{11}}\eta_{1t} = \eta_{1t}$. Thus, $\sum_{t=1}^{T} \nabla^2_{\theta_{11}\theta_{11}}\eta_{1t}\cdot\nabla_{\eta_1}L(y_t,\eta_t)\Big|_{\hat{\theta}_T} = \sum_{t=1}^{T} \nabla_{\theta_{11}}\eta_{1t}\cdot\nabla_{\eta_1}L(y_t,\eta_t)\Big|_{\hat{\theta}_T}$, which equals zero by the first-order conditions for the MLE for θ_{11}. So for θ_{11} the sample moment based on the first term in (2.2) will equal zero and therefore this part of the IM test will not test the mean of λ_t. Also, the first component in $\mathrm{vech}(\eta_{1t}^2X_tX_t') = \eta_{1t}^2$. Combining, the component of the IM test corresponding to the intercept parameter θ_{11} is a score test against the alternative that λ_t is random with mean η_{1t} and variance $T^{-1/2}\eta_{1t}^2{}'\alpha_{11}$, which coincides with the standard score test for heterogeneity in the exponential model and Poisson model (when $l = 2$). Clearly, this result hinges on the particular parameterization $\eta_{1t} = \exp(X_t'\theta_1)$.[3]

More generally, the IM test is a score test of a very particular specification of heterogeneity (of the type considered here) that will differ from that specified in standard heterogeneity tests.

---

[3] And even when $\eta_{1t} = \exp(X_t'\theta_1)$, the other non-intercept components of the IM test will not simplify in this way.

Appendix

Verification of (2.2). For simplicity suppress the subscript t. We have

$$\text{(A.1)}\qquad \nabla_{\theta_i} L(y,X,\theta) = \nabla_{\theta_i} L(y,\eta_1(X,\theta_1),\ldots,\eta_h(X,\theta_h)) = \nabla_{\theta_i}\eta_i(X,\theta_i)\cdot\nabla_{\eta_i} L(y,\eta)$$

and

$$\text{(A.2)}\qquad \nabla^2_{\theta_i\theta_j} L(y,X,\theta) = \nabla^2_{\theta_i\theta_j}\eta_i(X,\theta_i)\cdot\nabla_{\eta_i} L(y,\eta) + \nabla_{\theta_i}\eta_i(X,\theta_i)\cdot\nabla^2_{\eta_i\eta_j} L(y,\eta)\cdot(\nabla_{\theta_j}\eta_j(X,\theta_j))' .$$

The sum $\nabla^2_{\theta_i\theta_j} L(y,X,\theta) + \nabla_{\theta_i} L(y,X,\theta)\cdot(\nabla_{\theta_j} L(y,X,\theta))'$ yields (2.2).

Proof of lemma. For simplicity suppress the subscript t. We have:

$$\begin{aligned}\text{(A.3)}\quad f^*(y,\eta,\gamma,\Gamma) &= \int f(y,\lambda)p(\lambda,\eta,\gamma,\Gamma)\,d\lambda \\ &= \int \{f(y,\eta) + \nabla_\eta f(y,\eta)'(\lambda-\eta) \\ &\qquad + \tfrac{1}{2}(\lambda-\eta)'\nabla^2_\eta f(y,\eta)\cdot(\lambda-\eta) + O(T^{-1})\}\cdot p(\lambda,\eta,\gamma,\Gamma)\,d\lambda \\ &= f(y,\eta) + \nabla_\eta f(y,\eta)'\gamma + \tfrac{1}{2}\operatorname{tr}(\nabla^2_\eta f(y,\eta)\cdot\Gamma) + O(T^{-1}) \\ &= f(y,\eta) + \nabla_\eta f(y,\eta)'\gamma + \tfrac{1}{2}\operatorname{tr}(f(y,\eta)\cdot H(y,\eta)\cdot\Gamma) + O(T^{-1}) \\ &= f(y,\eta)\cdot[1 + \nabla_\eta L(y,\eta)'\gamma + \tfrac{1}{2}\operatorname{tr}(H(y,\eta)\cdot\Gamma) + O(T^{-1})] \\ &= f(y,\eta)\cdot\exp[\nabla_\eta L(y,\eta)'\gamma + \tfrac{1}{2}\operatorname{tr}(H(y,\eta)\cdot\Gamma)] + O(T^{-1}),\end{aligned}$$

where

$$\text{(A.4)}\qquad H(y,\eta) = \nabla^2_\eta L(y,\eta) + \nabla_\eta L(y,\eta)\cdot\nabla_\eta L(y,\eta)' ,$$

with $L(y,\eta) = \log(f(y,\eta))$. In (A.3), 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 mean-square error of λ given η and Γ; the fourth line uses the result that for any density under suitable regularity conditions $\nabla_\eta f(y,\eta) = f(y,\eta)\cdot\nabla_\eta L(y,\eta)$, since $\nabla_\eta L(y,\eta) = f(y,\eta)^{-1}\cdot\nabla_\eta f(y,\eta)$, and $\nabla^2_\eta f(y,\eta) = f(y,\eta)\cdot H(y,\eta)$, since $\nabla^2_\eta L(y,\eta) = f(y,\eta)^{-1}\cdot\nabla^2_\eta f(y,\eta)$ -

$\nabla_\eta L(y,\eta)\cdot\nabla_\eta L(y,\eta)'$; and the final line uses the approximation $\exp(x) = 1 + x$ for small x.

Proof of proposition 1. For $f^*(y)$ defined in (2.3) or (A.3), $\log f^*(y) = \nabla_\eta L(y,\eta)'\gamma + 0.5\cdot\mathrm{tr}(H(y,\eta)\cdot\Gamma)$, where $H(y,\eta)$ is defined in (A.4). The first term has derivative $\nabla_{\eta_i} L(y,\eta)$ w.r.t. γ_i, and the second term has derivative $0.5\cdot H_{ji}(y,\eta) = 0.5\cdot H_{ij}(y,\eta)$ w.r.t. Γ_{ij}. For γ_i and Γ_{ij} the particular functions of α_{ij} defined in (2.4) and (2.5), use of the chain rule differentiation immediately yields $\partial\log f^*(y_t)/\partial\alpha_{ij}$ equal to the expression within the expectation in (2.2).

References

Cameron, A.C., and P.K., Trivedi, 1986, Econometric Models based on Count Data: Comparisons and Applications of Some Estimators and Tests, *Journal of Applied Econometrics*, 1, 29-53.

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

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

Hildreth, C., and J.P.Houck, 1968, "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, 584-595.

Lee, L.-F, 1986, Specification Test For Poisson Regression Models, *International Economic Review* 27, 3, 689-706.

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