GENERALIZED METHODS OF MOMENTS

SPECIFICATION TESTING

Whitney K. Newey*
Princeton University

Econometric Research Program
Research Memorandum No. 306

June 1983

*The purpose of this paper is to provide a set of general results for specification tests based on moment conditions, which explicitly consider the power properties of these tests. These results include a theorem on the nonconsistency of moment specification tests, a result showing mutual asymptotic equivalence of maximal degree of freedom tests, and the derivation of optimal methods for testing a subset of moment conditions. The usefulness of these results is illustrated by several applications to particular tests.

*J. Hausman, D. McFadden, and K. Singleton provided useful comments on an earlier version of this paper.

I. Introduction

Most conclusions and predictions obtained by using econometric methods to summarize economic data are sensitive to model specification. The purpose of this paper is to provide a set of general results for specification tests based on moment conditions, including Hausman (1978) tests and Sargan (1958) and Hansen (1982) tests of overidentifying restrictions, which explicitly consider the power properties of specification tests. These results include a theorem on the general nonconsistency of moment specification tests, which emphasizes the importance of explicit power considerations, a result on mutual asymptotic equivalence of a certain class of specification tests, and the derivation of optimal methods for testing a subset of moment conditions.

The usefulness of the results of this paper are illustrated by several applications. Specification testing for a single equation of a simultaneous system is considered in some detail, in addition to applications which discuss the relationship of several specification tests in various contexts. Throughout the paper it is shown that results on asymptotic equivalence reduce to numerical equality for the special case of a linear in parameters model.

Most econometric estimators are formed by making use of certain functions of the data and parameters which have expectation zero when evaluated at the true parameter value. We refer to these functions as orthogonality condition functions. Most econometric estimators can be viewed as being obtained by minimizing a quadratic form in sample moments of the orthogonality condition functions. This class of estimators include maximum likelihood estimators, for which the elements of the score vector form orthogonality condition functions, and instrumental variables, for which cross-products of instruments and residuals form orthogonality condition functions. When more orthogonality condition functions than parameters are available, specification tests can be based on how close the sample moments of the orthogonality condition functions are to zero, when evaluated at the estimated parameter values. Tests based on the distribution of a linear combination of these sample moments will be referred to as generalized method of moments (GMM) specification tests in this paper.

In section two the general form of the test statistic which we discuss is presented and its asymptotic distribution is derived. It is shown that GMM specification tests are not consistent against general forms of misspecification. Section three discusses the relationship of Hausman (1978) specification tests and GMM tests. Section four presents some comparisons of the first order asymptotic power properties of different specification tests, including an asymptotic equivalence theorem for a class of maximal degree of freedom GMM tests and an optimal test for the validity of a subset of orthogonality conditions. Section five gives some applications of the general theory we develop and Section six presents some conclusions.

II. GMM Specification Tests

In order to discuss the formal properties of GMM specification tests we first develop some notation. Let $z = (z_1, z_2, \ldots)$ be a realization of a strictly stationary stochastic process, where z_t is an element of R^p.[1] Let the true parameter vector b_o be contained in a subset B of R^q, and let $g(z,b)$ be a vector-valued function from $B \times R^p$ to R^r. Define

$$g_T(b) = \frac{1}{T} \sum_{t=1}^{T} g(z_t, b) \tag{2.1}$$

A GMM estimator $\hat{b}_T$ of b_o will be assumed to be obtained as the solution to

$$\min_{b \in B} g_T(b)' W_T g_T(b) \tag{2.2}$$

where W_T is a rxr positive semi-definite matrix which depends on the data z. The estimator $\hat{b}_T$ is obtained by setting the sample moments $g_T(b)$ close to zero by minimizing the quadratic form $g_T(b)'W_T g_T(b)$. This class of estimators has been considered by Amemiya (1973), Burguete, Gallant and Souza (1982), and Hansen (1982) among others. If b_o, $g(z,b)$, and the stochastic process for z satisfy the property

$$E(g(z, b_o)) = 0 \tag{2.3}$$

so that the population moment $E(g_T(b))$ is equal to zero at the true parameter value b_o, then when appropriate regularity conditions, including identification, are satisfied the estimator b_T will be consistent for b_o. If specification error is present so that

$$E(g(z,b_o)) \neq 0, \tag{2.4}$$

it will often be the case that $\hat{b}_T$ is not consistent for b_o. When more orthogonality conditions than parameters are available, specification tests can be based on how close the sample moments are to zero when evaluated at the parameter estimates. The first order conditions for $\hat{b}_T$ are

$$g_{Tb}(\hat{b}_T)'W_T g_T(\hat{b}_T) = 0 \tag{2.5}$$

where $g_{Tb}(b) = \frac{\partial g_T}{\partial b}(b)$, so that $\hat{b}_T$ is obtained by setting linear combinations of the sample moments equal to zero. Specification tests can be based on how close other linear combinations of $g_T(\hat{b}_T)$ are to zero.

Let L_T be a sxr matrix which can depend on the data z. Then GMM tests are based on how close the s linear combinations $L_T g_T(\hat{b}_T)$ are to zero, after accounting for sampling error using asymptotic distribution theory. Let Q_T^- be a consistent estimator for Q^-, a generalized inverse of the asymptotic covariance matrix Q of $\sqrt{T}L_T g_T(\hat{b}_T)$. Then the form of the GMM specification test statistic is

$$m_T = T\, g_T(\hat{b}_T)'L_T'\, Q_T^-\, L_T g_T(\hat{b}_T).$$

The use of a generalized inverse allows for singularity of Q. Using standard arguments, and regularity conditions such as those which are presented in Hansen (1982) the asymptotic distribution of m_T can be shown to be chi-squared with degrees of freedom equal to rank (Q), when the model is correctly specified with $E(g(z,b_o)) = 0$. Of course there are many possible choices of L_T, and even many ways of forming $g(z,b)$ in most applications. In order to distinguish between different specification tests it is desirable to have some idea of the power of specification tests for detecting misspecification. In order to consider asymptotic power properties of specification tests we choose to consider a sequence of misspecification alternatives which will result in m_T having a noncentral chi-squared asymptotic distribution. We show how the usual asymptotic testing theory can be extended to allow for comparison of tests based on any GMM estimator, as well as comparison of maximum likelihood based tests as discussed in Hausman and Taylor (1980) and Holly (1982).

To treat local misspecification, we let c be a misspecification parameter which lies in R^u. For each n the distribution function of $(z_1,\ldots,z_n)$ will be specified as $F_n(z_1,\ldots,z_n,c)$, for each c. Where the expectation exists, define

$$h(b,c) = \int g(z,b)F_1(dz,c). \qquad (2.6)$$

For the purpose of exposition in the body of this paper we will assume that $h(b,c)$ is continuously differentiable in b and c, and that $g(z,b)$ is continuously differentiable in b. We will assume that at a point c_o,

$$h(b_o, c_o) = 0. \tag{2.7}$$

If $c = c_o$ the model is correctly specified, since the orthogonality conditions hold in the population.

We can now allow for local misspecification as follows. Let $c_T = c_o + \delta/\sqrt{T}$. We assume that for each T, $(z_1, \ldots, z_T)$ has a distribution function $F_T(z_1, \ldots, z_T; c_T)$ for each sample size T. Note that we are implicitly assuming that the assumed model is nested within the actual data generating process. However, our specification of local alternatives can allow for nonnested alternatives, in the same manner as is done in Ericsson (1983).

To derive the asymptotic distribution of m_T, some additional notation and assumptions are needed.

Assumption 2.1: The estimator $\hat{b}_T$ satisfies $\hat{b}_T \xrightarrow{p} b_o$, where b_o lies in the interior of B. Also, $h(b,c)$ exists and satisfies $h(b_o, c_o) = 0$.

Assumption 2.1 states that $\hat{b}_T$ is weakly consistent for b_o in the presence of local misspecification. We do not explicitly consider regularity conditions which are sufficient for the assumptions of this section to hold. One set of sufficient regularity conditions for the independent observations case is given in Newey (1983).

Assumption 2.2: The vector $g(z,b)$ is a measurable function on a measurable space Z, and for almost all $z \varepsilon Z$ a continuously differentiable function of b.

Assumption 2.3: The function $h(b,c)$ is continuously differentiable in b and c, $E\left(\frac{\partial g(z,b)}{\partial b}\right) = \frac{\partial h}{\partial b}(b,c)$, and $\frac{\partial g_T(b)}{\partial b}$ converges in probability to $\frac{\partial h}{\partial b}(b,c_o)$ uniformly in b on every compact subset of B.

Assumption 2.3 is implied by differentiability of density of z and of $g(z,b)$ along with dominance conditions (see Newey (1983)).

Define

$$H(b) = \frac{\partial h}{\partial b}(b,c_o) \ , \ H = H(b_o)$$

Assumption 2.4: The estimator b_T satisfies $\sqrt{T}[\partial g_T(\hat{b}_T)/\partial b']W_T g_T(\hat{b}_T) = o_p(1)$ for a sequence of matrices W_T satisfying plim $W_T = W$, W positive semi-definite and $H'WH$ nonsingular. Also, $L_T - L = o_p(1)$ for L with rank $(L) = s$.

Assumption 2.5: The random vector $Y_T = \sqrt{T}(g_T(b_o) - h(b_o,c_T))$ converges in distribution to a random variable $Y_o \sim N(0,V)$ where

$$V = \lim_{T\to\infty} T\, E(g_T(b_o)g_T(b_o)' - h(b_o,c_T)h(b_o,c_T)')$$

and V is nonsingular. Also $V_T \xrightarrow{p} V$.

The matrix V is the asymptotic covariance matrix of Y_T, and V_T is a consistent estimator of V. Methods of obtaining such a consistent V_T are outside the scope of this paper, but are considered in White (1980), Hansen (1982) and White and Domowitz (1982). Define the matrices

$$P_W = I - H(H'WH)^{-1}H'W,$$

$$H_T = \frac{\partial g_T}{\partial b}(\hat{b}_T) \quad , \quad P_{WT} = I - H_T(H_T'W_TH_T)^{-1}H_T'W_T$$

$$Q = LP_WVP_W'L' \quad , \quad Q_T = L_TP_{WT}V_TP_{WT}'L_T'$$

Assumption 2.6: The sequence of generalized inverses Q_T^- satisfies $Q_T^- \xrightarrow{p} Q^-$ where Q^- is a generalized inverse of Q.

Assumption 2.6 is required because Q may be singular. A sufficient condition for Assumption 2.6 to hold is that for all T a fixed generalized inverse (g-inverse) which is a continuous function of the elements of Q_T is chosen.[2]

It is useful to consider the special case which occurs when g(z,b) is linear in b.[3] Let

$$g(z,b) = G_1(z) - G_2(z)b \tag{2.8}$$

If we define $G_{1T} = (1/T) \sum_{t=1}^{T} G_1(z_t)$ and $G_{2T} = (1/T) \sum_{t=1}^{T} G_2(z_t)$ then, when $g(z,b)$ is linear in b, the estimator solving equation (2.2) is given by

$$\hat{b}_T = (G'_{2T} W_T G_{2T})^{-1} G'_{2T} W_T G_{1T}. \tag{2.9}$$

Define $\alpha = \partial h(b_o,c_o)/\partial c \cdot \delta$.

<u>Theorem 2.1</u>: If $c_T = c_o + \delta/\sqrt{T}$ and assumptions 2.1-2.6 are satisfied then

$$m_T = T\, g_T(\hat{b}_T)' L_T'\, Q_T^- L_T g_T(\hat{b}_T)$$

converges in distribution to a noncentral chi-squared distribution with degrees of freedom equal to rank(Q) and noncentrality parameter

$$\lambda^2 = \alpha' P_W' L' Q^- L P_W \alpha.$$

Also if m_T and m_T' correspond to two different choices of the sequence of g-inverses Q_T^- satisfying Assumption 2.6 then $m_T - m_T' = o_p(1)$. Further, if $g(z,b)$ is linear in b then m_T is invariant with respect to choice of g-inverse Q_T^-.

Proof: All proofs are given in the appendix.

An important property of GMM specification tests is that they are not consistent against general forms of misspecification. This inconsistency for some specific specification tests has been noted by Bierens (1982), who also discusses methods of forming consistent specification tests for regression models. It is our purpose to show that this inconsistency is a fundamental phenomenon and is related to identification of parameters under misspecification.

We can in fact show that it is a general property of GMM specification tests that they are inconsistent against a subset of the alternative space which has dimension equal to the dimension of the alternative space minus the degrees of freedom of the test. To state this result it is necessary to consider nonlocal misspecification. Let z be a realization from a stochastic process with a fixed c not necessarily equal to c_o. Let

$$\text{plim } L_T = L(c) \text{ , plim } W_T = W(c) \text{ and plim } V_T = V(c)$$

We impose the following different assumptions.

<u>Assumption 2.7</u>: The function g(z,b) is measurable in z and twice continuously differentiable in b. The function h(b,c) exists, and is twice continuously differentiable in b and c for each b in B and c in C, where C is an open subset of R^u. Further, for each c in C, $g_T(b)$

and its first and second partial derivatives converge in probability to $h(b,c)$ and its first and second partial derivatives, respectively, uniformly in b.

Assumption 2.8: For each c in C, $h(b,c)'W(c)h(b,c)$ has a unique minimum for $b = b(c)$ in the interior of B. Also, B is compact.

Assumption 2.9: $L(c)$ and $V(c)$ are once continuously differentiable functions of c in C.

Assumption 2.10: For each c in C, $\sqrt{T}(L_T - L(c)) = O_p(1)$ and $\sqrt{T}(W_T - W(c)) = O_p(1)$.

Assumption 2.11: For each c in C, $\sqrt{T}(g_T(b(c)) - h(b(c),c)) = O_p(1)$

and $\sqrt{T}(g_{Tb}(b(c)) - \frac{\partial h}{\partial b}(b(c),c)) = O_p(1)$.

Define $L_o = L(c_o)$, $W_o = W(c_o)$, $H_o = \frac{\partial h}{\partial b}(b_o,c_o)$, $K_o = \frac{\partial h}{\partial c}(b_o,c_o)$ and $V_o = V(c_o)$.

Assumption 2.12: The sxu matrix

$$L_o = (I - H_o(H_o'W_oH_o)^{-1}H_o'W_o)K_o$$

has rank s. Also $H_o'W_oH_o$ and V_o are nonsingular.

Note that assumption 2.12 implies that $s > u$.

Theorem 2.2: If Assumptions 2.7-2.12 are satisfied, then there is an open set N in R^u containing c_o such that the set of c in N satisfying $m_T = 0_p(1)$ is a u-s dimensional C^1 sub manifold of N.

Note that Assumption 2.12 implies that the rank of

$$Q_o = L_o(I - H_o(H_o'W_oH_o)^{-1}H_o'W_o)V_o(I - W_oH_o(H_o'W_oH_o)^{-1}H_o')L_o'$$

is s, which is also the degrees of freedom of the GMM specification test, which shows that the set of c values for which the GMM specification test does not reject with probability approaching one for every fixed significance level has dimension equal to the difference of the dimension of the alternative space and the degrees of freedom of the test.

The nonconsistency of GMM specification tests can be explained in terms of parametric identification. If the model is misspecified, so that the true value $\bar{c} \neq c_o$, then

$$E(g(z,b_o) - h(b_o,\bar{c})) = 0$$

by the definition of $h(b,c)$. Define a new orthogonality condition function

III. Hausman and GMM Specification Tests

Before discussing different GMM specification tests, it is important to clarify the relationship between these tests and Hausman tests. We first consider Hausman tests in the GMM estimation framework. Let $\tilde{b}_T$ and $\bar{b}_T$ be two GMM estimators such that $\tilde{b}_T$ is obtained as the solution to

$$\min_b g_T(b)'A_T g_T(b) \tag{3.1}$$

and $\bar{b}_T$ from

$$\min_b g_T(b)'C_T g_T(b) \tag{3.2}$$

To assumptions 2.1-2.6 we add the following assumptions

Assumption 3.1: $\bar{b}_T \xrightarrow{p} b_o$ and $\tilde{b}_T \xrightarrow{p} b_o$.

Assumption 3.2: $A_T \xrightarrow{p} A$ with A positive semi-definite and H'AH is non-singular. $C_T \xrightarrow{p} C$ with C positive semi-definite and H'CH is non-singular.

The estimators $\bar{b}_T$ and $\tilde{b}_T$ satisfy $\sqrt{T}\, g_{Tb}(\bar{b}_T)'\, C_T\, g_T(\bar{b}_T) = o_p(1)$ and $\sqrt{T}\, g_{Tb}(\tilde{b}_T)'A_T g_T(\tilde{b}_T) = o_p(1)$.

Let

$$q_T = \tilde{b}_T - \bar{b}_T$$

and

$$M = (H'CH)^{-1}H'CVCH(H'CH)^{-1} + (H'AH)^{-1}H'AVAH(H'AH)^{-1} \qquad (3.3)$$
$$- (H'CH)H'CVAH(H'AH)^{-1} - (H'AH)^{-1}H'AVCH(H'CH)^{-1}$$

The matrix M will be the asymptotic covariance matrix of q_T. A consistent, positive semi-definite estimator of M_T can be obtained from equation (3.3) by replacing V and V_T, A by A_T, C by C_T and H by e.g.,

$$\tilde{H}_T = g_{Tb}(\tilde{b}_T). \qquad (3.4)$$

Assumption 3.3: The sequence of g-inverses M_T^- is chosen so that plim $M_T^- = M^-$.

As with Assumption 2.6, Assumption 3.3 is required due to possible singularity of M.

Now define the test statistic

$$h_T = T\, q_T'\, M_T^-\, q_T$$

Theorem 3.1: If Assumptions 2.2-2.5 and 3.1-3.3 are satisfied then h_T converges in distribution to a noncentral chi-squared distibution with degrees of freedom d_h= rank (M) and noncentrality parameter

$$\lambda_h^2 = \alpha'[AH(H'AH)^{-1} - CH(H'CH)^{-1}]\, M^-[(H'AH)^{-1}H'A - (H'CH)^{-1}H'C]\alpha$$

Also, if h_T and h_T' correspond to different choices of M_T^- satisfying Assumption 3.3, then $h_T - h_T' = o_p(1)$. Further, if $g(z,b)$ is linear in b then h_T does not depend on the choice of g-inverse M_T^-.

It should be emphasized that Theorem 3.1 gives the asymptotic distribution under local misspecification of most of the Hausman tests which have been proposed in the literature.[1,2] The orthogonality condition functions $g(z,b)$ and the choice of A and C give lots of latitude for putting Hausman tests in the form of Theorem 3.1. Note also that the form of misspecification we consider includes various forms of correlation of variables with residuals and finitely parameterized likelihood function misspecification.

An attraction of the Hausman test, as presented in Hausman (1978), is its computational simplicity. Note that if $C = V^{-1}$, then

$$M = (H'AH)^{-1}H'AVAH(H'AH)^{-1} - (H'V^{-1}H)^{-1}, \tag{3.4}$$

which is the difference of the asymptotic covariance matrices of $\tilde{b}_T$ and $\bar{b}_T$. As shown in Hansen (1982), when $C = V^{-1}$, $\bar{b}_T$ is asymptotically efficient relative to $\tilde{b}_T$, for any choice of A satisfying assumptions 3.1 and 3.2. Therefore, it is sufficient for M to have the simple difference form, as discussed in Hausman (1978), that one estimator used in forming q_T corresponds to the efficient choice of weighting matrix W, for a given set of orthogonality condition functions.[3] This result implies that all of the specification tests discussed in Hausman (1978) have the simple matrix difference form even if the disturbances are not normally distributed.

There is a simple, asymptotic relationship between Hausman tests and GMM specification tests. This relationship follows from a one-step theorem for GMM estimators. Where the inverse exists, define

$$\dot{b}_T = \tilde{b}_T - (\tilde{H}_T' C_T \tilde{H}_T)^{-1} \tilde{H}_T' C_T g_T(\tilde{b}_T). \qquad (3.5)$$

Theorem 3.2: If Assumptions 2.2-2.5 and 3.1-3.3 are satisfied, then $\sqrt{T}(\bar{b}_T - \dot{b}_T) = o_p(1)$. Further, if $g(z,b)$ is linear in b then $\bar{b}_T = \dot{b}_T$.

Theorem 3.2 is the appropriate generalization to GMM estimators of the well known one-step theorems for maximum likelihood and nonlinear least squares, and holds when the model is locally misspecified.

Theorem 3.2 implies that

$$\sqrt{T}q_T = \sqrt{T}(\dot{b}_T - \tilde{b}_T) + o_p(1) \tag{3.6}$$

$$= -(\tilde{H}_T'C_T\tilde{H}_T)^{-1}\tilde{H}_T'C_T\sqrt{T}g_T(\tilde{b}_T) + o_p(1)$$

so that by non-singularity of H'CH, a Hausman test based on the difference $q_T = \tilde{b}_T - \bar{b}_T$ is asymptotically equivalent to a GMM specification test with

$$W = A \; , \; L = H'C. \tag{3.7}$$

and is equal to the GMM test with $W_T = A_T$ and $L_T = H_T'C_T$ if $g(z,b)$ is linear in b. Similarly, starting at $\bar{b}_T$ and taking one step in the direction of $\tilde{b}_T$ it is evident that a Hausman test based on $q = \tilde{b}_T - \bar{b}_T$ is also asymptotically equivalent to a GMM specification test with

$$W = C \; , \; L = H'A. \tag{3.8}$$

The equivalence of Hausman tests and tests based on moment conditions has also been discussed in Ruud (1982) and White (1982). This view of Hausman tests as GMM tests helps to facilitate first order asymptotic comparisons, as will be illustrated in section four.

Theorem 3.2 also brings out the fact that Hausman tests are inconsistent against general forms of misspecification. An immediate consequence of the first order asymptotic equivalence of $\dot{b}_T$ and $\tilde{b}_T$ is that the noncentrality parameter λ_h^2 of Theorem 3.1 will be zero on a subset of δ values in R^u which is an u-d_h dimensional linear subspace of R^u. A result exactly analogous to Theorem 2.2 can be shown to imply that the set of alternatives for which a Hausman test fails has dimension equal to the dimension of the alternative space minus the degrees of freedom of the test.

IV. Comparing Local Power of GMM Specification Tests

The local power of different GMM specification tests can be compared by comparing their respective noncentral chi-squared distributions. The tail probability of a noncentral chi-squared distribution is increasing in the noncentrality parameter and decreasing in the numbers of degrees of freedom. The following two results give a convenient method of determining degrees of freedom of specification tests when the asymptotic covariance matrix, V, of $\sqrt{T}g_T(b_o)$ is nonsingular. Let R(A) denote the rank of the matrix A.

Proposition 4.1. If V and H'WH are nonsingular then

$$R(Q) = R([WH, L']) - q$$

where q = dim(b) and H, W and Q have been previously defined.

Corollary 4.2: If V, H'AH and H'CH are nonsingular then

$$R(M) = R([AH, CH]) - q.$$

Recall that Q is the asymptotic covariance of the sample moments $L_T g_T(\hat{b}_T)$, M is the asymptotic covariance matrix of the difference of two estimators $q_T = \tilde{b}_T - \bar{b}_T$, and that q = dim(b). We will use these results to obtain the degrees of freedom of particular tests and these results should prove to be useful in other applications.

Somewhat surprisingly, there is a general class of asymptotically equivalent GMM tests. Consider two different choices of GMM specification tests statistics m_{1T} and m_{2T} corresponding to different choices of the linear combination matrix L and the weighting matrix W.

Proposition 4.3: If Assumptions 2.1-2.6 are satisfied by both m_{1T} and m_{2T}, and if the degrees of freedom of the asymptotic distribution of m_{1T} and m_{2T} equal r-q, then $m_{1T} - m_{2T} = o_p(1)$. Further, if g(z,b) is linear in b then $m_{1T} = m_{2T}$.

To interpret this result, note that the asymptotic covariance matrix Q equals $LP_W VP_W' L'$. Furthermore, $R(Q) \leq r-q$, since P_W is an idempotent matrix and

$$R(P_W) = \text{trace}(P_W) = \text{trace}(I) - \text{trace}(H(H'WH)^{-1}H'W) = r-q.$$

Since the degrees of freedom of a GMM tests is R(Q), it follows that the degrees of freedom of a GMM test is less than or equal to r-q. A restatement of Proposition 4.3 is that any GMM test with the maximum number of degrees of freedom, r-q, is asymptotically equivalent to any other GMM test with degrees of freedom r-q, and numerical equality holds if g(z,b) is linear in b.[1]

As a benchmark for comparison, it is useful to know the maximum value of the noncentrality parameter.

Proposition 4.4: The noncentrality parameter λ^2 satisfies

$$\lambda^{*2} = \alpha' P_W'(P_W V P_W')^- P_W \alpha \geq \lambda^2$$

Further, λ^{*2} does not depend on W.

This maximum value of the noncentrality parameter is attained by the GMM test with degrees of freedom r-q. Choose $L_T = I$ for all T, where I is an rxr identity matrix. For a given weighting matrix W, the GMM test with $L_T = I$ has degrees of freedom

$$R(Q) = R([WH, I]) - q = r-q,$$

and by Proposition 4.1, and by Theorem 2.1 has noncentrality parameter

$$\lambda^2 = \alpha' P_W' I'(I P_W V P_W' I')^- I P_W \alpha = \lambda^{*2}.$$

As discussed in Section two, GMM specification tests are consistent only if the parameters of the model are identified when misspecification is present. Also, comparison of different specification tests depends on the form of the alternative considered. A form of alternative which maintains identification when the model is misspecified, and which is interesting in applications is given by specifying

$$E(g(z, b_0)) = \bar{c}$$

where $\bar{c}$ is restricted to have at least q components zero. This restriction means that when the model is misspecified, at least q orthogonality conditions remain valid. This form of alternative leads directly to GMM specification tests for subsets of orthogonality conditions.

For local power purposes, misspecification which results in some orthogonality conditions being contaminated and other orthogonality conditions remaining valid corresponds to a partitioning of α as

$$\alpha = (0,\alpha_2')',$$

where, α_2 is kx1 vector, with $r - k > q$. Let $U = V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1}$. Choosing $W = V^{-1}$ and defining $P = P_W$, the maximum value for the noncentrality parameter is given by

$$\lambda^{*2} = \alpha'P'(PVP')^{-}P\alpha = \alpha'P'V^{-1}P\alpha = \alpha'U\alpha = \alpha_2'U_{22}\alpha_2$$

where as noted in Hansen (1982), V^{-1} is a g-inverse of PVP' and U is partitioned conformably with α,

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}.$$

Also, partition H, V, and V_T conformably with $\alpha = (0,\alpha_2')'$,

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \quad V_T = \begin{bmatrix} V_{11T} & V_{12T} \\ V_{21T} & V_{22T} \end{bmatrix}.$$

The maximum value of λ^{*2} is attained by any GMM specification test with degrees of freedom r-q. If q = r - k, so that the parameters are exactly identified under misspecification, the only consistent tests will be GMM tests with degrees of freedom k = r - q, which are all asymptotically equivalent by Proposition 4.3. When q < r - k, so that the parameters are overidentified under misspecification, a consistent test with higher local power than the r-q degrees of freedom test can be obtained. One such optimal test can be obtained as follows.

<u>Proposition 4.5</u>. If Assumptions 2.1-2.6 are satisfied with $W_T = V_T^{-1}$, and

$L_T = [-V_{21T}V_{11T}^{-1}, I_k]$, $\alpha = (0, \alpha_2')$, and H_1 has rank q, then $\bar{m}_T$ converges in distribution to a noncentral chi-squared distribution with k degrees of freedom and noncentrality parameter $\lambda^{*2} = \alpha_2' U_{22} \alpha_2$. Also, U_{22} is non-singular.

The test statistic $\bar{m}_T$ has the following form. For

$Q_T = L_T(V_T - H_T(H_T'V_TV_TH_T)^{-1}H_T')L_T'$ and $g_T(\hat{b}_T) = (g_{T1}(\hat{b}_T)', g_{T2}(\hat{b}_T)')'$

partitioned conformably with α,

$$\bar{m}_T = [g_{T2}(\hat{b}_T)' - g_{T1}(\hat{b}_T)'V_{11T}^{-1}V_{12T}]Q_T^{-1} \quad (4.1)$$

$$[g_{T2}(\hat{b}_T) - V_{21T}V_{11T}^{-1}g_{T1}(\hat{b}_T)],$$

where $\hat{b}_T$ is the optimal GMM estimator which is obtained by solving

$$\min_b g_T(b)'V_T^{-1} g_T(b).$$

The linear combination matrix L_T, used in forming $\bar{m}_T$, forms from $g_T(\hat{b}_T)$ the sample counterpart to the population residuals of the regressions of the elements of the limit of $\sqrt{T}\, g_{T2}(b_o)$ on the limit of $\sqrt{T}\, g_{T1}(b_o)$ (i.e. $\bar{m}_T$ partials out $g_{T1}(\hat{b}_T)$).

Another specification test which is relevant when some orthogonality conditions are contaminated and others are not is a Hausman test based on the difference of b_T and the GMM estimator $\tilde{b}_T$ which is obtained by solving

$$\min_b g_{1T}(b)'V_{11T}^{-1} \; g_{1T}(b). \tag{4.2}$$

The estimator $\tilde{b}_T$ is the optimal GMM estimator which uses all orthogonality conditions which remain uncontaminated under misspecification. From the discussion of Hausman tests we know that the asymptotic covariance matrix of $q_T = \tilde{b}_T - \hat{b}_T$ will be

$$\bar{M} = (H_1'V_{11}^{-1}H_1)^{-1} - (H'V^{-1}H)^{-1}.$$

Let the Hausman specification test statistic be given by

$$\overline{h_T} = Tq_T'\bar{M}_T^{-}q_T.$$

<u>Proposition 4.6</u>: If Assumptions 2.2-2.5 and 3.1-3.3 are satisfied for the Hausman test based on $q_T = \tilde{b}_T - \bar{b}_T$ then $\overline{h_T}$ converges in distribution

to a noncentral chi-squared distribution with degrees of freedom

$$d_h^* = R\left(\begin{bmatrix} V_{11} & H_1 \\ V_{21} & H_2 \end{bmatrix}\right) + k-r \qquad (4.3)$$

and non-centrality parameter

$$\lambda_h^{*2} = \alpha' V^{-1} H(H'V^{-1}H)^{-1}[(H_1'V_{11}^{-1}H_1)^{-1} - (H'V^{-1}H)^{-1}]^{-}(H'V^{-1}H)^{-1}H'V^{-1}\alpha .$$

Note that as long as the matrix in equation (4.3) has maximal rank equal to min $(r + q - k, r)$, then $d_h^* = \min(q,k)$.

It is possible to compare the two specification tests based on $\bar{h}_T$ and $\bar{m}_T$. A comparison of these two tests is analogous to the discussion in Holly (1982) and Hausman and Taylor (1980). In particular the following proposition is true.

Proposition 4.7: If $\bar{h}_T$ and $\bar{m}_T$ have the same degrees of freedom then $\bar{h}_T - \bar{m}_T = o_p(1)$. Further, if $\bar{h}_T$ and $\bar{m}_T$ have the same degrees of freedom and $g(z,b)$ is linear in b, then $\bar{m}_T = \bar{h}_T$. If $\bar{h}_T$ and $\bar{m}_T$ have different degrees of freedom, then the asymptotic power curves of $\bar{h}_T$ and $\bar{m}_T$ will cross.

Our results on power comparisons can be summarized as follows. For a given set of orthogonality condition functions $g(z,b)$, all GMM specification tests with the maximum number of degrees of freedom r-q are asymptotically equivalent. If the parameters are overidentified when misspecification is present, then there are consistent GMM tests which are more powerful than the r-q degrees of freedom test.

V. Applications

Our first application of the theoretical results we have obtained is to the estimation framework presented by Hansen (1982). Let $\tilde{m}_T = T\, g_T(\hat{b}_T)'(P_{WT}V_TP'_{WT})^- g_T(\hat{b}_T)$. Use of this test statistic was suggested by Hansen as a general test of model misspecification. Theorem 2.2 implies that this test is inconsistent against general forms of misspecification. When misspecification is parameterized by plim $g_T(b_o)$ the implicit null hypothesis is that $\delta = \lim \sqrt{T}\, E(g_T(b_o))$ belongs to the linear space spanned by the columns of $E[\partial g_T(b_o)/\partial b]$. Also, Proposition 4.3 implies that the local power of this test against any form of misspecification is invariant with respect to the choice of $W = \text{plim}\, W_T$ satisfying our assumptions so that the asymptotic properties of $\tilde{m}_T$ are independent of the GMM estimator $\hat{b}_T$. For $g(z,b)$ linear in b, the test statistics are invariant with respect to W_T.

The test statistic $\tilde{m}_T$ can also be interpreted as a Hausman test for $r < 2q$. Let A_T be any rxr matrix satisfying plim $A_T = A$ with rank$[AH,WH] = r$. Then for $\tilde{b}_T$ obtained from equation (3.1) Corollary 4.2 implies the Hausman test based on the difference $\hat{b}_T - \tilde{b}_T$ has degrees of freedom r-q. By Proposition 4.3 and the equivalence of this Hausman test to a GMM test, this Hausman test is asymptotically equivalent to the test statistic $\tilde{m}_T$.

Finally, in time series estimation there are situations where optimal tests of a subset of orthogonality conditions are useful. In the context of estimation of rational expectations models, the orthogonality condition functions are often obtained as cross products of disturbances and random variables belonging to an agents information set. The tests we have presented can therefore be used to form optimal tests of whether particular sets of random variables belong to the information sets. These tests can also be used for testing covariance restrictions, as discussed in Newey (1983).

Our second application concerns tests of overidentifying restrictions in the linear simultaneous equations system. Note, first, that our results have an interesting implication for linear system specification tests. It follows from Proposition 4.3 and our discussion of Hausman tests that the system test of overidentifying restrictions based on the Gallant and Jorgenson (1979) criteria for three-stage least squares (3SLS) (i.e. the Hansen(1982) minimum chi-square test for the linear simultaneous equations system) is numerically equal to the specification test based on the difference of the two-stage least squares (2SLS) and 3SLS estimators suggested by Hausman (1978), when the two tests have the same degrees of freedom, and the same estimate of the covariance matrix of the disturbances is used to form both statistics. Our results also have implications for specification tests of a single equation, which we will consider in some detail.

Without loss of generality, let the first equation of a simultaneous system be written in regression form as

$$y_1 = Y_1\beta + Z_1\delta + u = X_1 b + u, \tag{5.1}$$

where exclusion restrictions and the usual normalization have been imposed, $X_1 = [Y_1, Z_1]$ and $b = (\beta', \gamma')'$. The vector y_1 is a Tx1 vector of observations on the left-hand side endogenous variable, Y_1 is a Txp matrix of observations on the included right-hand side endogenous variables, Z_1 is a Txs matrix of observations on the included predetermined variables and u is a Tx1 vector of disturbances. The px1 vector β gives the endogenous variable coefficients, the sx1 vector γ gives the predetermined variable coefficients and $p+s = q$ is the number of parameters to be estimated. Let $Z = [Z_1, Z_2]$ be the Txr matrix of observations on the predetermined variables of the system and $Y = [Y_1, Y_2]$ be the Tx(M-1) matrix of endogenous variables of the system except for y_1, where the system contains a total of M equations.

We will assume that there is no autocorrelation or heteroscedasticity, so that when $b=b_o$, the true parameter vector,

$$\frac{Z'u}{\sqrt{T}} \xrightarrow{d} N(0,V) \ ; \ V = \sigma^2 \ \text{plim}\left(\frac{Z'Z}{T}\right); \ \sigma^2 = E[u_t^2 | Z_t]. \tag{5.2}$$

The subscript t indexes the $t^{\underline{th}}$ observation, $t=1,\ldots,T$. We will consider specification tests based on the 2SLS estimator b_T, which solves

$$\min_{b} u'Z(Z'Z)^{-1}Z'u. \tag{5.3}$$

Let $\hat{u} = y_1 - X_1\hat{b}_T, \hat{\sigma}^2 = \hat{u}'\hat{u}/T$ and $P_z = Z(Z'Z)^{-1}Z'$. Throughout the following discussion we will take $V_T = \hat{\sigma}^2 Z'Z/T$.

The general minimum chi-square test statistic for 2SLS is given by

$$\tilde{m}_T = \frac{\hat{u}'Z}{\sqrt{T}} V_T^{-1} \frac{Z'\hat{u}}{\sqrt{T}} = T\hat{u}'P_z\hat{u}/\hat{u}'\hat{u}$$

see Sargan (1958). As noted by Hausman (1983), $\tilde{m}_T = T\,\tilde{R}^2$ where $\tilde{R}^2$ is the r-squared from a regression of $\hat{u}$ on Z.

There are two particular types of departures from correct specification which may be of concern for a single simultaneous equation. One is that certain variables have been wrongly excluded from the equation, and the other is that certain instrumental variables (corresponding to columns of Z) may be correlated with u_o. If the equation remains overidentified under misspecification, then the theoretical results of the previous section indicate that more powerful specification tests against these particular alternatives than that based on $\tilde{m}_T$ are available.

We consider first the case where misspecification takes the form of wrongly excluded variables. Suppose that the correct model is given by

$$y_1 = Y_1\beta + Z_1\gamma + WC + \varepsilon \tag{5.5}$$

where W consists of $\ell < r-q$ columns of $[Y_2, Z_2]$. Following Burgete, Gallant and Souza (1982) we can define a gradient test of H_0: C=0 using the criterion function of the minimization problem (5.3) which is used to obtain the 2SLS estimator.[1] For ε defined implicitly in equation (5.5) we have

$$\partial\varepsilon'Z(Z'Z)^{-1}Z'\varepsilon/\partial C = -W'Z(Z'Z)^{-1}Z'\varepsilon \tag{5.6}$$

so that a gradient test which uses $\hat{b}_T$ is based on the asymptotic distribution of $W'Z(Z'Z)^{-1}Z'\hat{u}/\sqrt{T}$ and is thus a GMM test with linear combination matrix $L_T = W'Z(Z'Z)^{-1}$.

Then

$$[V_T^{-1}H_T, L_T'] = [\hat{\sigma}^2(Z'Z)^{-1}Z'X_1, (Z'Z)^{-1}Z'W] \tag{5.7}$$

so that by Proposition 4.1 the degrees of freedom of this test will be ℓ so long as plim $Z'[X_1, W]T$ has rank $q+\ell$. Define $\hat{X}_1 = P_Z X_1$ and $\hat{W} = P_Z W$. In terms of our previous notation, $H_T = -Z'X_1/T$ so that

$$V_T - H_T(H_T'V_T^{-1}H_T)^{-1}H_T = \hat{\sigma}^2 Z'M_{\hat{X}_1}Z/T \tag{5.8}$$

where $M_{\hat{X}_1} = I_T - \hat{X}_1(\hat{X}_1'\hat{X}_1)^{-1}\hat{X}_1'$. The GMM (gradient) test statistic is then given by

$$m_T = \hat{u}'ZL_T'[L_TZ'M_{\hat{X}_1}ZL_T']^{-1}L_TZ'\hat{u}\cdot(1/\hat{\sigma}^2) \quad (5.9)$$
$$= T\hat{u}'\hat{W}(\hat{W}'M_{\hat{X}_1}\hat{W})^{-1}\hat{W}'\hat{u}/\hat{u}'\hat{u},$$

where the non-singularity of the matrix $\hat{W}'M_{\hat{X}_1}\hat{W}$ for large T follows from the degrees of freedom of m_T being equal to ℓ. Note that it follows from the normal equations $\hat{X}_1'\hat{u} = 0$ that $m_T = TR^2$ where R^2 is the r-squared from a regression of $\hat{u}$ on $[\hat{X}_1, \hat{W}]$.

While we have not explicitly considered the power properties of tests of parametric hypotheses such as H_o: C = 0, the following direct argument shows that m_T gives an optimal GMM test. Let C_T satisfy $\lim \sqrt{T}C_T = \delta$. Then by $u_o = WC_T + \varepsilon$, an appropriate central limit theorem implies $Z'u_o\sqrt{T} \xrightarrow{d} N(\text{plim}(Z'W/T)\delta, V)$ so that

$$\lambda^{*2} = \delta'\text{plim}(W'Z/T)[V_T^{-1} - V_T^{-1}H_T(H_T'V_T^{-1}H_T)^{-1}H_T'V_T^{-1}](Z'W/T)\delta = \delta'\text{plim}(1/T\hat{\sigma}^2)(\hat{W}'M_{\hat{X}_1}\hat{W})\delta \quad (5.10)$$

Further, since $L_TP_TZ' = W'Z(Z'Z)^{-1}[I_K - Z'X_1(X_1'X_1)^{-1}X_1'Z(Z'Z)^{-1}]Z' = \hat{W}M_{\hat{X}_1}$, the non-centrality parameter of m_T is given by

$$\lambda^2 = \delta'\ \text{plim}[(1/T\sigma^2)W'ZP_T'L_T'(W'M_{X_1}W)^{-1}L_TP_TZ'W]\delta \quad (5.11)$$
$$= \delta'\text{plim}[(1/T\sigma^2)\hat{W}'M_{\hat{X}_1}\hat{W}(\hat{W}'M_{\hat{X}_1}\hat{W})^{-1}\hat{W}'M_{\hat{X}_1}\hat{W}]\delta = \lambda^{*2}.$$

The second particular form of misspecification which is of interest involves instrumental variable contamination. For this form of misspecification we can use the optimal GMM test of Proposition 4.5. Suppose that $Z = [Z^1, Z^2]$ where Z^2 is a Txk vector, $k < r-q$, which is correlated with u_o when misspecification is present and Z^1 is a Tx(r-k) vector of predetermined variables which remain uncorrelated with u_o. Let $H_1 = -\operatorname{plim} Z^{1\prime}X_1/T$ have rank q, so that the parameters b_o remain identified under misspecification. Then from Proposition 4.5 this optimal GMM test has $L_T = [-Z^{2\prime}Z^1(Z^{1\prime}Z^1)^{-1}, I_k]$. Note that $ZL_T' = \tilde{V} = (I_T - Z^1(Z^{1\prime}Z^1)^{-1}Z^{1\prime})Z^2$ is the Txk matrix of residuals from a regression of the columns of Z^2 on Z^1. The test statistic is given by

$$\bar{m}_T = \hat{u}'ZL_T'(L_TZ'M_{\hat{X}_1}ZL_T')^{-1}L_TZ'\hat{u}/\hat{\sigma}^2 \tag{5.12}$$

$$= T\,\hat{u}'\tilde{V}(\tilde{V}'M_{\hat{X}_1}\tilde{V})^{-1}\tilde{V}'\hat{u}/\hat{u}'\hat{u} = T\bar{R}^2$$

where $\bar{R}^2$ is the r-squared from the regression of $\hat{u}$ on $[\hat{X}_1, \tilde{V}]$.

Another test for instrument contamination, which is particularly useful when primary interest centers on the parameter vector b, is a Hausman test based on the difference of $\hat{b}_T$ and the 2SLS estimator of b using only Z^1 as instrumental variables. This Hausman test statistic has been considered by Hausman and Taylor (1980) and Spencer and Berk (1981). By Proposition 4.6 this Hausman test has degrees of freedom given by

$$\bar{d}_h = \text{rank}(\text{plim } Z'[Z^1, X_1]/T) + k - r \qquad (5.13)$$

As discussed in Hausman and Taylor (1980), the rank of the matrix plim $Z'[Z^1, X_1]/T$ is equal to $\min(r, q+r-k-n)$, where n is the number of common columns of Z^1 and Z_1. Then $\bar{d}_h = \min(k, q-n)$. By Theorem 3.2 and linearity in b the Hausman test statistic is equal to a GMM test with $L_T = [X_1'Z^1(Z^{1\prime}Z^1)^{-1}, 0]$. Let $\hat{\hat{X}}_1' = X_1'Z^1(Z^{1\prime}Z^1)^{-1}Z^{1\prime}$, and let S' be a $\bar{d}_h$ x q selection matrix such that $S'\hat{\hat{X}}_1'M_{\hat{X}_1}\hat{\hat{X}}_1S'$ is non-singular. Then by Lemma A2 of the appendix and the invariance of the GMM statistic with respect to g-inverse

$$\begin{aligned}\bar{h}_T &= \hat{u}'ZL_T'(L_TZ'M_{\hat{X}_1}ZL_T')^- L_TZ'\hat{u}/\hat{\sigma}^2 \qquad (5.14)\\ &= \hat{u}'\hat{\hat{X}}_1(\hat{\hat{X}}_1'M_{\hat{X}_1}\hat{\hat{X}}_1)^-\hat{\hat{X}}_1'\hat{u}/\hat{\sigma}^2\\ &= T\hat{u}'\hat{\hat{X}}_1S(S'\hat{\hat{X}}_1'M_{\hat{X}_1}\hat{\hat{X}}_1S)^{-1}S'\hat{\hat{X}}_1'\hat{u}/\hat{u}'\hat{u}\\ &= T\bar{R}_h^2\end{aligned}$$

where $\bar{R}_h^2$ is the R-squared from a regression of $\hat{u}$ on $[\hat{X}_1, \hat{\hat{X}}_1S]$.[2]

Note that Proposition 4.7 implies that if $\bar{d}_h = k$, then $\bar{h}_T = \bar{m}_T$. If the set of predetermined variables being tested for contamination, Z^2, is a subset of the predetermined variables included in the equation, Z_1, $k + n < q$, so that $\bar{d}_h = k$, and for the important special case of testing for the predetermined status of included variables, $\bar{h}_T = \bar{m}_T$.

When a full set of overidentifying restrictions is being tested then each of these test statistics is identical. That is, if $\ell = k = \bar{d}_h = r-q$, then Proposition 4.3 implies that $\tilde{m}_T = m_T = \bar{m}_T = \bar{h}_T$. Note that these equalities hold independently of the particular set of overidentifying coefficient restrictions being tested or the particular subset of instrumental variables being tested for contamination. When the misspecification of interest is omitted variables or endogeneity of an instrumental variable, and the equation is overidentified under misspecification, then the appropriate test gives an optimal GMM test. However, even when a particular form of misspecification occurs, the statistic $\tilde{m}_T$ retains a certain kind of robustness. Its noncentrality parameter is at least as large as that of the individual statistics, no matter what form the misspecification takes. Asymptotic power loss from use of $\tilde{m}_T$ rather than a specific statistic will result from $\tilde{m}_T$ having larger degrees of freedom.

VI. Conclusions.

We have presented results for a class of specification tests, which we have referred to as generalized method of moments specification tests, which include Hausman (1978) tests. Due to lack of identification under general misspecification, GMM specification tests are not consistent. We have shown that all maximal degree of freedom GMM tests are mutually equivalent, asymptotically. When specific forms of misspecification are considered such that the model parameters are identified under misspecification, consistent GMM tests can be compared on the basis of their local power. Overidentification under misspecification leads to specification tests which are locally more powerful than the maximal degree of freedom test of overidentifying restrictions.

The results of this paper also show that specification tests can be found wherever there are more orthogonality functions than parameters to be estimated. In many econometric models, there are an infinite number of such orthogonality condition functions available. It is often the case in econometric models that there is an nxl vector $e(z,b)$ function of z and b such that, if the model is correctly specified, the conditional expectation $E[e(z_t,b_o)|I_t]$ satisfies

$$E[e(z_t,b_o)|I_t] = 0, \tag{6.1}$$

where I_t is a conditioning set. In the estimation of rational expectations modls $e(z_t,b_o)$ is often a vector of forecast errors and I_t is the information set available to an agent at time t. Then for any rxn random variable $w(I_T)$ satisfying $E|w(I_t)| < +\infty$ and $E|w(I_t)e(z_t,b_o)| < +\infty$ the law of iterated expectations (Chung (1974)) implies

$$E[w(I_t)e(z_t,b_o)] = E[w(I_t)E[e(z_t,b_o)|I_t]] = 0. \qquad (6.2)$$

Therefore we can use as orthogonality condition functions $g(z_t,b) = w(I_t)e(z_t,b)$.

There are very many ways of picking the $w(I_t)$ random variables to form specification tests, which illustrates that in most econometric models there will be many ways of forming a specification test. It is therefore important to pick a test statistic which is appropriate for a particular application. In this paper we have given results which allow an econometrician to distinguish among different specification tests based on classical power considerations, for a particular set of moment condition functions. In Newey (1983) we give methods of picking the optimal $w(I_t)$ to form an optimal set of orthogonality condition functions, where again the optimality criteria employed are classical power considerations.

Appendix

We first give several lemmas which are useful in the proofs that follow.

Lemma A1: (Rao(1973), 1.b.5, (vi),a): For a matrix A, $A(A'A)^{-}A'A = A$ and $A'A(A'A)^{-}A' = A'$ for any choice of g-inverse.

Lemma A2: (Rao and Mitra (1971) Lemma 2.2.5(b)): For conformable matrices A and B, if $R(ABA') = R(B)$, then $A'(ABA')^{-}A$ is a g-inverse of B for any choice of $(ABA')^{-}$.

Lemma A3: (Rao and Mitra (1971) Lemma 2.2.6(g)): For conformable matrices A and B, if $R(ABA') = R(A)$ then $A'(ABA')^{-}A$ is invariant for any choice of g-inverse.

Lemma A4: For conformable matrices A and B, $A(A'A)^{-}A'$ and $A(A'A)^{-}A' - AB(B'A'AB)^{-}B'A'$ are idempotent.

Proof: Idempotency of $A(A'A)^{-}A'$ follows immediately from Lemma A1. Also, by Lemma A1

$$
\begin{aligned}
&(A(A'A)^{-}A' - AB(B'A'AB)^{-}B'A')^2 \\
&= A(A'A)^{-}A' - A(A'A)^{-}A'AB(B'A'AB)^{-}B'A' \\
&- AB(B'A'AB)^{-}B'A'A(A'A)^{-}A' \\
&+ AB(B'A'AB)^{-}B'A'AB(B'A'AB)^{-}B'A' \\
&= A(A'A)^{-}A' - AB(B'A'AB)^{-}B'A'.
\end{aligned}
$$

For a matrix A, let N(A) be the null space of A and C(A) the column space of A.

Lemma A5: Let A be a $k \times \ell$ matrix, B a $\ell \times m$ matrix and C a $\ell \times n$ matrix. If the columns of C form a basis for N(A) and R(B) = m, then, R(AB) =

R([C,B]) - n.

Proof: For x in N(AB), let y = Bx. Then y is an element of N(A), so that by C a basis for N(A) there is a unique z such that Cz = y = Bx, which implies [C,B] [-z',x']' = 0. Similarly, suppose [C,B] [-z',x']' = 0. Then Cz = Bx implies ABx = ACz = 0. Therefore N(AB) is isomorphic to N([C,B]) and so dim N(AB) = dim N([C,B]). By Lancaster (1969) Theorem 1.16.2 R(AB) = m - dimN(AB) and R([C,B]) = m + n - dimN([C,B]), so that R(AB) - m = R[C,B] - m - n or R(AB) = R[C,B] - n.

Lemma A6: For comformable matrices A and B, if B is positive definite, then $R(A'(ABA')^{-}A) = R(A)$.

Proof: We know $R(A) \geq R(A'(ABA')^{-}A)$. By the definition of a g-inverse

$$AB(A'(ABA')^{-}A)BA' = ABA'$$

so that $R(A'(ABA')^{-}A) \geq R(ABA') = R(A)$, where the last equality follows by the positive definiteness of B.

Proof of Theorem 2.1: For notational convenience, we will suppress the z argument. Since b_o lies in the interior of B and plim $\hat{b}_T = b_o$, the first condition of assumption 1.4 implies

$$g_{Tb}(\hat{b}_T)'W_T\sqrt{T}\, g_T(\hat{b}_T) = o_p(1). \tag{A.1}$$

Without changing notation, we consider a sequence of random variables tail equivalent to $\hat{b}_T$ which lie in an open convex neighborhood of b_o which is contained in the interior of B. Apply a mean value expansion to obtain

$$\sqrt{T}\, g_T(\hat{b}_T) = \sqrt{T}\, g_T(b_o) + g_{Tb}(\tilde{b}_T)\sqrt{T}\,(\hat{b}_T - b_o) \tag{A.2}$$

where $|\tilde{b}_T - b_o| \leq |\hat{b}_T - b_o|$ and $\tilde{b}_T$ actually differs from row to row of $g_{Tb}(\tilde{b}_T)$. Since plim $\hat{b}_T = b_o$, plim $\tilde{b}_T = b_o$. Equations (A.1) and (A.2) imply

$$g_{Tb}(\hat{b}_T)'W_T g_{Tb}(\tilde{b}_T)\sqrt{T}(\hat{b}_T - b_o) = -\, g_{Tb}(\hat{b}_T)'W_T\sqrt{T}\, g_T(b_o) + o_p(1) \tag{A.3}$$

By Assumption 2.3, plim $\hat{b}_T = b_o$ and plim $\tilde{b}_T = b_o$, plim $g_{Tb}(\hat{b}_T)$ = plim $g_{Tb}(\tilde{b}_T)$ = H, so that

$$\text{plim } g_{Tb}(\hat{b}_T)'W_T g_{Tb}(\tilde{b}_T) = H'WH \tag{A.4}$$

and

$$\text{plim } g_{Tb}(\hat{b}_T)'W_T = H'W \tag{A.5}$$

by the usual rules for probability limits of sums and products of random variables. By Assumption 2.5 and the definition of h(b,c),

$$\sqrt{T}\ g_T(b_o) = Y_T + \sqrt{T}\ h(b_o,c_T) = 0_p(1) + \sqrt{T}\ h(b_o,c_T). \qquad (A.6)$$

Take a mean value expansion of $h(b_o, c_T)$ around c_o to obtain

$$\sqrt{T}\ h(b_o,c_T) = \sqrt{T}\ h(b_o,c_o) + \frac{\partial h}{\partial C}(b_o,\tilde{c}_T)\sqrt{T}\ (c_T - c_o) \qquad (A.7)$$

$$= \frac{\partial h}{\partial c}(b_o,c_T)\delta + O(1)$$

where $|\tilde{c}_T - c_o| \leq |c_T - c_o|$. By Assumption 2.3 and $\lim c_T = c_o$, $\lim \sqrt{T}\ h(b_o,c_T) = \lim \frac{\partial h}{\partial c}(b_o,c_T)\delta = \alpha$. Then by equations (A.6) and (A.7)

$$\sqrt{T}\ g_T(b_o) = Y_T + \alpha + o(1) = 0_p(1). \qquad (A.8)$$

Then by equations (A.8), (A.3) and (A.5)

$$g_{Tb}(\hat{b}_T)'W\ g_{Tb}(\tilde{b}_T)\sqrt{T}\ (\hat{b}_T - b_o) = -\ H'W(Y_T + \alpha) + o_p(1). \qquad (A.9)$$

Then since H'WH is non-singular by Assumption 2.4, $(g_{Tb}(\hat{b}_T)'W_T g_{Tb}(\tilde{b}_T))^{-1}$ exists with probability approaching one, and by equations (A.9) and (A.4)

$$\sqrt{T}\ (\hat{b}_T - b_o) = -\ (H'WH)H'W(Y_T + \alpha) + o_p(1) = 0_p(1). \qquad (A.10)$$

Now, expand $\sqrt{T}\ g_T(\hat{b}_T)$ around b_o to find

$$\sqrt{T}\, g_T(\hat{b}_T) = \sqrt{T}\, g_T(b_o) + g_{Tb}(\tilde{b}_T)\, \sqrt{T}\, (\hat{b}_T - b_o) \tag{A.11}$$

$$= Y_T + \alpha + o_p(1) + H\sqrt{T}(b_T - b_o) + o_p(1) = P_w(Y_T + \alpha) + o_p(1) = O_p(1)$$

where $|\tilde{b}_T - b_o| \leq |\hat{b}_T - b_o|$, and the last three equalities follow from equation (A.10) and the arguments leading to it. Then by plim $L_T = L$ and $\sqrt{T}\, g_T(\hat{b}_T) = O_p(1)$,

$$L_T\sqrt{T}\, g_T(\hat{b}_T) = LP_w(Y_T + \alpha) + o_p(1) \xrightarrow{d} N(\alpha, Q) \tag{A.12}$$

by Assumption 2.5. By Assumption 2.6, equation (A.12) and $\sqrt{T}\, L_T g_T(\hat{b}_T) = O_p(1)$,

$$m_T = Tg_T(\hat{b}_T)'L_T'Q_T^-L_Tg_T(\hat{b}_T) + o_p(1) \tag{A.13}$$

$$= (Y_T + \alpha)'P_w'L'Q^-LP_w(Y_T + \alpha) + o_p(1)$$

from which follows the fact that m_T converges in distribution to a noncentral chi-squared with degrees of freedom rank (Q) and noncentrality parameter λ^2. The asymptotic equivalence of m_T and m_T' follows from Lemma A3 and Lemma A6, which imply $P_w'L'Q^-LP_w$ is invariant with respect to g-inverse of Q.

To show that when g(z,b) is linear in b, m_T is invariant with respect to choice of g-inverse, note that

$$g_T(\hat{b}_T) = G_{1T} - G_{2T}\hat{b}_T = [I_r - G_{2T}(G_{2T}'W_TG_{2T})^{-1}G_{2T}'W_T]G_{1T} \tag{A.14}$$

$$= P_{WT}G_{1T},$$

so that

$$m_T = TG_{1T}'P_{WT}'L_T'(L_TP_{WT}V_TP_{WT}'L_T')^{-}L_TP_{WT}G_{1T}, \tag{A.15}$$

and invariance of m_T follows by the same argument as asymptotic equivalence of m_T and m_T'.

Proof of Theorem 2.2: Let $J(c) = L(c)h(b(c),c)$. Note that by $b_o = b(c_o)$ and $h(b_o,c_o) = 0$, $J(c_o) = L_o0 = 0$, and $\frac{\partial J}{\partial c}(c_o) = L_o[H_o \frac{\partial b}{\partial c}(c_o) + K_o]$. Let $f(b,c) = (1/2)h(b,c)'W(c)h(b,c)$. Then by the definition of $b(c)$, and $b(c)$ in the interior of B, $b(c)$ solves

$$\frac{\partial h}{\partial b}(b(c),c)'W(c)h(b(c),c) = 0. \tag{A.16}$$

By the implicit function theorem, $h(b_o,c_o) = 0$ and $H_o'WH_o$ nonsingular, equation (A.16) implies

$$\frac{\partial b}{\partial c}(c_o) = -(H_o'W_oH_o)^{-1}H_o'W_oK_o. \tag{A.17}$$

Then by (A.16),

$$\frac{\partial J}{\partial c}(c_o) = L_o[I - H_o(H_o'W_oH_o)^{-1}H_o'W_o]K_o.$$

By Assumption 2.12, $\frac{\partial J}{\partial c_o}$ is a sxu matrix of rank s. The vector function $J(c)$ is continuously differentiable in an open neighborhood N' of c_o by Assumption 1.7 and 1.9 and the implicit function theorem applied to equation (A.16). By J(c) continuous on N', J(c) has rank s on for all c in an open neighborhood $N'' \subset N'$. Then by the implicit function theorem (e.g., Hirsch (1976) Theorem A.9) the set of c in N' such that J(c) = 0 is a C^1, u-s dimensional submanifold of N. It now suffices to show that for c in N', if J(c) = 0, then $m_T = O_p(1)$.

Let $f_T(b) = 1/2\ g_T(b)'W_Tg_T(b)$. By Assumption 2.7, $f_T(b)$ converges in probability to f(b,c) uniformly in b. Assumption 2.8 and a convergence in probability version of Amemiya (1973) Lemma 3 implies that there exists a measureable $\hat{b}_T$ solving

$$\min_{b \in B} \ f_T(b) \tag{A.18}$$

and satisfying plim $\hat{b}_T = b(c)$. Since b(c) lies in the interior of B, $\hat{b}_T$ satisfies

$$g_{Tb}(\hat{b}_T)'\ W_T \sqrt{T}\ g_T(\hat{b}_T) = o_p(1). \tag{A.19}$$

Expanding $g_T(\hat{b}_T)$ in a mean value expansion (as in the proof of Theorem 2.1) equation (A.19) implies

$$\begin{aligned} &g_{Tb}(\hat{b}_T)'W_Tg_{Tb}(\hat{b}_T)\sqrt{T}\ (\bar{b}_T - b(c)) + \\ &\quad g_{Tb}(\hat{b}_T)W_T \sqrt{T}\ g_T(b(c)) = o_p(1) \end{aligned} \tag{A.20}$$

with $|\bar{b}_T - b(c)| < |\hat{b}_T - b(c)|$. Define $H(c) = \frac{\partial h}{\partial b}(b(c),c)$. By adding and subtracting appropriate terms, using equation (A.16)

$$g_{Tb}(\hat{b}_T)'W_T\sqrt{T}\ g_T(b(c)) = g_{Tb}(\hat{b}_T)'W_T\sqrt{T}\ (g_T(b(c)) - h(b(c),c)) \quad (A.21)$$

$$+ g_{Tb}(\hat{b}_T)\sqrt{T}(W_T - W(c))\ h(b(c),c)$$

$$+ \sqrt{T}\ [g_{Tb}(\hat{b}_T) - H(c)]'W(c)h(b(c),c)$$

By Assumption 2.7, and plim $\hat{b}_T = b(c)$, plim $g_{Tb}(\hat{b}_T) = H(c)$. Then by Assumptions 2.10 and 2.11 the first two terms after the equality in equation (A.21) are $o_p(1)$. By Assumption 2.7 we can apply a mean-value expansion to the last term to obtain

$$\sqrt{T}[g_{Tb}(\hat{b}_T) - H(c)] = \sqrt{T}[g_{Tb}(b(c)) - H(c)] \quad (A.22)$$

$$+ \sum_{j=1}^{q} \frac{\partial g_{Tb}}{\partial b_j}(\bar{b}_T)(\hat{b}_{Tj} - b(c)_j)\sqrt{T}$$

with $|\bar{b}_T - b(c)| < |\hat{b}_T - b(c)|$. Then by Assumption 2.10 and equations (A.21) and (A.22)

$$g_{Tb}(\hat{b}_T)'W_T g_{Tb}(\hat{b}_T)\sqrt{T}(\hat{b}_T - b(c)) +$$

$$\sum_{j=1}^{q} \frac{\partial g_{Tb}}{\partial b_j}(\bar{b}_T)W(c)h(b(c),c)\sqrt{T}(\hat{b}_{Tj} - b(c)_j) = o_p(1).$$

By Assumption 2.7 and plim $\hat{b}_T = b(c)$, plim $g_{Tb}(\hat{b}_T)$ = plim $g_{Tb}(\bar{b}_T)$ = $H(c)$, and plim $\frac{\partial g_{Tb}}{\partial b_j}(\bar{b}_T) = \frac{\partial}{\partial b_j}[\frac{\partial h}{\partial b}(b(c),c)]$. By continuity in c,

$H_o'W_oH_o$ non-singular and $h(b_o,c_o) = 0$, it follows that there is $N \subset N''$ such that for c in N,

$$\sqrt{T}(\hat{b}_T - b(c)) = O_p(1). \qquad (A.23)$$

Now, to show that $L_T\sqrt{T}g_T(\hat{b}_T) = O_p(1)$ if $J(c) = 0$ for c in N, the mean value expansion of $\sqrt{T}\ g_T(\hat{b}_T)$ implies, using $J(c) = L(c)h(b(c),c) = 0$,

$$L_T\sqrt{T}\ g_T(\hat{b}_T) = L_T g_{Tb}(\tilde{b}_T)\sqrt{T}\ (b_T - b_o) \qquad (A.24)$$
$$+ \sqrt{T}[L_T - L(c)]g_T(b(c))$$
$$+ L(c)\sqrt{T}[g_T(b(c)) - h(b(c),c)].$$

The first and second terms are $O_p(1)$ by equation (A.23), its proof, Assumption 2.10, and Assumption 2.7 which implies plim $g_T(b(c)) = h(b(c),c)$ so that $g_T(b(c)) = O_p(1)$. By Assumption 2.11, the second term is also $O_p(1)$.

Proof of Proposition 2.3: For a positive matrix B, and conformable A, $N(A) \subset N(A'(ABA')^- A)$ is obvious. Suppose $A'(ABA')^- Ax = 0$. Then premultiplying by AB and letting $C = B^{1/2}$

$$0 = ABA'(ABA')^- Ax = ACCA'(ACCA')^- ACC^{-1}x$$
$$= ACC^{-1}x = Ax$$

where the third equality follows from Lemma A1. The conclusion then follows, upon noting that $\alpha = \partial h(b_o,c_o)/\partial c \cdot \delta$, from taking $A = LP_W$ and

$B = V$.

Proof of Theorem 3.1: The assumptions of Theorem 1.1 are satisfied for $\tilde{b}_T$ and $\bar{b}_T$, so that equation (A.10) of the proof of Theorem 1.1 implies

$$\sqrt{T}(\tilde{b}_T - b_o) = -(H'A'H)^{-1}HA(Y_T + \alpha) + o_p(1) \qquad (A.25)$$

and

$$\sqrt{T}(\bar{b}_T - b_o) = -(H'CH)^{-1}H'C(Y_T + \alpha) + o_p(1). \qquad (A.26)$$

Subtraction yields

$$\sqrt{T}\, q_T = [(H'CH)^{-1}H'C-(H'AH)^{-1}H'A](Y_T + \alpha) + o_p(1). \qquad (A.27)$$

Let $D = (H'CH)^{-1}H'C - (H'AH)^{-1}H'A$. Then $\sqrt{T}\, q_T = D(Y_T + \alpha) + o_p(1)$, so that

$$\sqrt{T}\, q_T \xrightarrow{d} D(Y_o + \alpha) \sim N(D\alpha, DVD'). \qquad (A.28)$$

Note that $M = DVD'$, so that by $\sqrt{T}\, q_T = O_p(1)$,

$$\begin{aligned} h_T &= Tq_T'(M_T^- - M^-)q_T + Tq_T'M^-q_T \\ &= Tq_T'M^-q_T + o_p(1) \end{aligned} \qquad (A.29)$$

and the asymptotic distribution result follows. To see that $h_T - h_T' = o_p(1)$, note that $D'(DVD')^-D$ is invariant with respect to g-inverse and the conclusion follows by the same argument as used in the proof of Theorem 2.1.

To see that h_T is invariant with respect to the g-inverse for the linear case, note that

$$\tilde{b}_T - \bar{b}_T = - D_T G_{1T} \tag{A.30}$$

for $D_T = (G'_{2T} C_T G_{2T})^{-1} G'_{2T} C_T - (G'_{2T} A_T G_{2T})^{-1} G'_{2T} A_T$

Further,

$$h_T = T(\tilde{b}_T - \bar{b}_T)'(D_T V_T D'_T)^{-}(\tilde{b}_T - \bar{b}_T) \tag{A.31}$$

$$= TG'_{1T} D'_T (D_T V_T D'_T)^{-} D_T G_{1T}$$

and the invariance of h_T follows from the invariance of $D'_T(D_T V_T D'_T)^{-} D_T$, which is implied by Lemma A3 and Lemma A6.

Proof of Theorem 3.2: From the proof of Theorem 1.4, it follows that

$$\sqrt{T}(\bar{b}_T - b_o) = - (H'CH)^{-1} H'C(Y_T + \alpha) + o_p(1) \tag{A.32}$$

so that

$$\sqrt{T}(\bar{b}_T - \dot{b}_T) = \sqrt{T}(\bar{b}_T - b_o) - \sqrt{T}(\dot{b}_T - b_o) \tag{A.33}$$

$$= -(H'CH)^{-1}H'C(Y_T + \alpha) - \sqrt{T}(\tilde{b}_T - b_o) + (\tilde{H}_T'C_T\tilde{H}_T)^{-1}\tilde{H}_T'C_T\sqrt{T}g_T(\tilde{b}_T) + o_p(1)$$

$$= -[I - (\tilde{H}_T'C_T\tilde{H}_T)^{-1}\tilde{H}_T'C_Tg_{Tb}(\ddot{b}_T)]\sqrt{T}(\tilde{b}_T - b_o)$$

$$+ [(\tilde{H}_T'C_T\tilde{H}_T)^{-1}\tilde{H}_T'C_T - (H'CH)^{-1}H'C](Y_T + \alpha) + o_p(1),$$

where $|\ddot{b}_T - b_o| < |\tilde{b}_T - b_o|$ and the last equality follows by expanding $g_T(\tilde{b}_T)$ around b_o. Upon noting that $Y_T + \alpha = O_p(1)$ and $\sqrt{T}(\tilde{b}_T - b_o) = O_p(1)$, it follows that $\sqrt{T}(\bar{b}_T - \dot{b}_T)$ is $o_p(1)$, since plim $(\tilde{H}_T'C_T\tilde{H}_T)^{-1} = (H'CH)^{-1}$, plim $\tilde{H}_T'C_Tg_{Tb}(\ddot{b}_T) = H'CH$ and plim $\tilde{H}_T'C_T = HC$.

Further, if $g(z,b)$ is linear in b, then

$$\dot{b}_T = \tilde{b}_T - (G_{2T}'C_TG_{2T})^{-1}(-G_{2T}')C_T(G_{1T} - G_{2T}\tilde{b}_T) \tag{A.34}$$

$$= \tilde{b}_T - (G_{2T}'C_TG_{2T})^{-1}G_{2T}'C_TG_{2T}\tilde{b}_T + (G_{2T}'C_TG_{2T})^{-1}G_{2T}'C_TG_{1T}$$

$$= \bar{b}_T$$

Proof of Proposition 4.1: The proof consists of showing that

$$R(Q) = R([WH,L']) - q. \tag{A.35}$$

Since $Q = LP_wVP_w'L'$ and V is positive definite, $R(Q) = R(P_w'L')$. Since $P_w' = I - WH(H'WH)^{-1}H'$ is idempotent, $R(P_w') = r-q$. Since $H'WH$ is nonsingular, $q = R(H'WH) \leq R(WH)$ implies $R(WH) = q$, so that the q columns of WH are linearly independent. By Lancaster (1969) Theorem 1.6.2 $\dim N(P_W') = q$, and since $P_W'WH = WH - WH(H'WH)^{-1}H'WH = 0$ the columns of WH form a basis for $N(P_W')$. Then by $R(L') = s$, Lemma A5 implies

$$R(P_W'L') = R([WH,L']) - q.$$

Proof of Corollary 4.2: Theorem 3.2 implies a GMM test with W = A, L = H'C is asymptotically equivalent to a Hausman test based on $q_T = \tilde{b}_T - \bar{b}_T$. Applying Proposition 4.1,

$$R(M) = R([AH,L']) - q = R([AH,CH]) - q$$

Proof of Proposition 4.3: From the proof of Theorem 2.1 it follows that

$$m_T = (Y_T + \alpha)'P_W'L'Q^-LP_W(Y_T + \alpha) + o_p(1).$$

Therefore it suffices to show that if $R(Q) = r-q$, then

$$P_W'L'Q^-LP_W = U = V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1}. \tag{A.36}$$

We know that $R(P_WVP_W') = r-q$. If $R(LP_WVP_W'L') = r-q$, then by Lemma A2,

$$L'(LP_WVP_W'L')^-L = (P_WVP_W')^-$$

so that by Lemma A3

$$P_w'(P_wVP_w')^-P_w = P_w'L'(LP_wVP_w'L')\overline{L}P_w.$$

Then it suffices to show

$$U - P_w'(P_wVP_w')^-P_w = 0.$$

Let F be a symmetric square root of V, with $F^2 = V$. Then

$$U - P_w'(P_wVP_w')^-P_w = F^{-1}[I-F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1} \tag{A.37}$$

$$- FP_w'P_wFFP_w') - P'F]F^{-1}.$$

Now $F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1}$ and $FP_w'(P_wFFP_w')^-P_wF$ are idempotent by Lemma A4. Further, $[FP_w'(P_wFFP_w')^-P_wF]\ F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1}] = 0$ by $P_wH = 0$, so that $I-F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1} - FP_w'(P_wFFP_w')^-P_wF = G$ is idempotent. Therefore

$$R(G) = \text{trace}(G) - r\text{-}q\text{-trace}(P_wVP_w'(P_wVP_w)^-) = 0$$ by Rao (1973).

For g(z,b) linear in b, equation (A.15) implies that we can replace $Y_T + \alpha$ by G_{1T} and P_W, V, L by P_{WT}, V_T, L_T respectively in the above argument to give numerical equality of m_{1T} and m_{2T}.

Proof of Lemma 1.4: The difference $\lambda^{*2} - \lambda^2$ satisfies

$$\lambda^{*2} - \lambda^2 = \alpha'F^{-1}[FP_W'(P_WF^2P_W')^-P_WF - FP_W'L'(LP_WF^2P_W'L')LP_WF]F^{-1}\alpha \tag{A.38}$$

which is non-negative, since the matrix in square brackets is idempotent by Lemma A4.

Proof of Proposition 4.5: Note that $[0,I_K]V^{-1} = (V^{-1})_{22}L$, so that by nonsingularity of $(V^{-1})_{22}$ we can take $L = [0,I_K]V^{-1}$. Then $LP = [0,I_K]U$ and by $UVU = U$, $Q = [0,I_K]U[0,I_K]' = U_{22}$. Then the degrees of freedom of this test are $\text{rank}(Q) = \text{rank}(U_{22}) = k$, by Proposition 4.1. Also

$$P'L'Q^{-}LP = U[0,I_K]'U_{22}^{-1}[0,I_K]U \qquad (A.39)$$

so that $\lambda^2 = \alpha'P'L'Q^{-}LP\alpha = \alpha_2'U_{22}U_{22}^{-1}U_{22}\alpha = \lambda^{*2}$.

Proof of Proposition 4.6: By Theorem 3.2, this Hausman test is asymptotically equivalent to a GMM test with $W = V^{-1}$ and

$L = H'\begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ and numerically equal if $g(z,b)$ is linear. By Proposition 4.1, the degrees of freedom of this test is

$$\begin{aligned} &R([V^{-1}H,\ L']) - q \\ &= R([H,\ VL']) - q = R\left(\begin{bmatrix} H_1 & H_1 \\ H_2 & V_{21}V_{11}^{-1}H_1 \end{bmatrix}\right) - q \\ &= R\left(\begin{bmatrix} H_1 & 0 \\ H_2 & H_2-V_{21}V_{11}^{-1}H_1 \end{bmatrix}\right) - q = R(H_2-V_{21}V_{11}^{-1}H_1) \end{aligned}$$

by $R(H_1) = q$. Since $H_2 - V_{21}V_{11}^{-1}H_1 = [-V_{21}V_{11}^{-1}\ \ I_k]H$, $R([-V_{21}V_{11}^{-1}, I_k]) = k$, so that $\dim N([-V_{21}V_{11}^{-1}\ I_k]) = r-k$, and

$[-V_{21}V_{11}^{-1}, I_k] \begin{bmatrix} V_{11} \\ V_{21} \end{bmatrix} = 0$ with $[V_{11}' \; V_{21}']'$ of rank r-k by V non singular, Lemma A6 implies

$$R(H_2 - V_{21}V_{11}^{-1} H_1) = R(\begin{bmatrix} V_{11} & H_1 \\ V_{21} & H_2 \end{bmatrix}) - (r-k).$$

Let $A = \begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$. Then $H_1'V_{11}^{-1}H_1 = H'AH = H'AVAH$ so that $M = (H_1'V_{11}^{-1}H_1)^{-1} - (H'V^{-1}H)^{-1}$. The form of the noncentrality parameter given then follows by Theorem 3.1.

Proof of Proposition 4.7: Let $W = V^{-1}$, so that $P_W = P$, and let $L = [H_1'V_{11}^{-1}, 0]$ be the linear combination matrix for the GMM version of this Hausman test. Define the kxq matrix, $B = H_2 - V_{21}V_{11}^{-1}H_1$. Straightforward calculation shows that $LV-H' = [0, -B']$. Then by $P = VU$ and $H'U = 0$, $LP = LVU = (LV-H')U = -B'[0,I_K]U$ and $Q = B'[0,I_K]U[0,I_K]'B$. It follows that

$$P'L'Q^{-}LP = U[0,I_K]'B(B'[0,I_K]U[0,I_K]'B)^{-}B'[0,I_K]U \qquad (A.40)$$

From the proof of Proposition 4.6 the degrees of freedom of the Hausman test is $d_h^* = \text{rank}(B)$. If $d_h^* = k$, then by nonsingularity of U_{22} and Lemma A2, $B(B'U_{22}B)^{-}B' = U_{22}^{-1}$. Then equation (A.39) gives

$$P'L'Q^{-}LP = U[0,I_K]'U_{22}^{-1}[0,I_K]U \tag{A.41}$$

Then asymptotic equivalence of m_T and h_T for $d_h = k$ follows from equations (A.39) and (A.41).

The numerical equivalence of $\bar{h}_T$ and m_T for rank(B) = k follows from the exact same argument, replacing V by V_T, H by $H_T = - G_{2T}$, B by $B_T = H_{T2} - V_{T21}V_{T11}^{-1}H_{T1}$, P by $P_T = I - H_T(H_T'V_T^{-1}H_T)^{-1}H_T'V_T^{-1}$, U by $U_T = V_T^{-1}P_T$ and $Y_T + \alpha$ by G_{1T}. Then $\bar{h}_T = G_{1T}'U_T[0,I_k]'U_{T22}^{-1}[0,I_k]U_TG_{1T} = \bar{m}_T$.

Now, if rank(B) ≠ k, rank(B) < k, so that the degrees of freedom of the Hausman test are less than the degrees of freedom of m_T. It follows that there are α_2 values such that $\alpha_2 \neq 0$ and $\lambda_h^{*2} = 0$ implying m_T has higher local power. If $\alpha_2 = B\gamma$, so that $\alpha = [0,I_k]'B\gamma$, equation (A.40) implies

$$\lambda_h^{*2} = \gamma'B'[0,I_k]U[0,I_k]'B(B'[0,I_k]U[0,I_k]'B)^{-}B'[0,I_k]U[0,I_k]'B\gamma$$

$$= \gamma'B'U_{22}B\gamma = \lambda^{*2},$$

so that $d_h^* < k$ implies the Hausman test has higher local power.

Footnotes

Section II

[1] The stationarity assumption is not essential. Research in progress indicates that an asymptotic testing theory is available even when moment matrices do not converge. Thus our stochastic assumptions could be weakened and the methods of, e.g., Domowitz and White (1982), could be used. The stationarity assumption simplifies notation, without changing any of the essential results.

[2] One method of obtaining the sequence Q_T^- is to form $Q_T^- = S_T'(S_T Q_T S'_T)^{-1} S_T$, where plim $S_T = S$, rank(Q) = rank(SQS'), and SQS is nonsingular. Then $Q^- = S'(SQS')^{-1}S$ follows from Lemma A2. Holly and Monfort (1982) have presented methods of forming a generalized inverse when required for a maximum likelihood Hausman test.

[3] The most important case where g(z,b) is linear in b, is when a linear equation is estimated by instrumental variables. Such an example is discussed in section five.

Section III

[1] Specification tests which use different orthogonality condition functions, such as the test based on the difference of two weighted least squares estimators suggested in Domowitz and White (1982), can be accomodated by stacking all the orthogonality condition functions into one vector and specifying that A_T and C_T have all zeros in certain rows and columns.

[2] The expressions $-(H'AH)^{-1}H'A\alpha$ and $-(H'CH)^{-1}H'C\alpha$ are the directional derivatives of plim $\tilde{b}_T$ and plim $\bar{b}_T$ with respect to c in the direction δ, evaluated at c_o. Thus, λ_h^2 is a quadratic form in the difference of the derivatives of the asymptotic bias of $\tilde{b}_T$ and $\bar{b}_T$.

[3] An alternative, more complicated, derivation of the matrix difference form of the covariance matrix of q_T can be obtained using the asymptotic Rao-Blackwell theorem of Hausman (1978, Lemma 2.1). Using Lemma 2.1 of Spencer and Berk (1981) it can be shown that for any q dimensional nonsingular matrix J, $b_T = J\tilde{b}_T + (I-J)\bar{b}_T$ is also a GMM estimator of b_o using the same orthogonality condition functions.

Section IV

[1] We have implicitly assumed that the same estimator V_T of V is used in forming each test statistic. Each of our results on numerical equality depend on this assumption.

Section V

[1] These tests will be based on the 2SLS estimator, which is consistent only when the null hypothesis of correct specification is true, so that the tests will be gradient (or Lagrange multiplier) tests.

[2] This derivation makes precise the method by which the Hausman (1978) test of Spencer and Berk (1981) can be obtained in an expanded regression framework. The form of this test statistic is that of Lagrange multiplier test for the inclusion of $\hat{\hat{X}}_1 S$ in a regression with right-hand side variables $\hat{X}_1$ and residual vector $\hat{u}$. Note that the statistic $\bar{m}_T$ is numerically equal to $\bar{h}_T$ in the Spencer and Berk (1981) case, and explicitly avoids having to choose a selection matrix S. An alternative method for testing for the exogeneity of a subset of included variables is given by Holly (1982).

Holly, A. and A. Montfort, 1982, Some Useful Equivalence Properties of Hausman's Test, mimeo.

Jennrich, R. I., 1969, Asymptotic Properties of Non-Linear Least Squares Estimators, Annals of Mathematical Statistics, 40, 633-643.

Lancaster, P., 1969, Theory of Matrices. New York: Academic Press.

Newey, W. K., 1983, Specification Testing and Estimation Using a Generalized Method of Moments. Unpublished Ph.D. Thesis, MIT.

Rao, C. R., 1973, Linear Statistical Inference and its Applications New York: John Wiley and Sons.

Rao, C. R. and S. K. Mitra, 1971, Generalized Inverse of Matrices and its Applications. New York: John Wiley and Sons.

Ruud, P., 1982, A Score of Consistency, mimeo, U.C. Berkeley.

Sargan, J.D., 1958, Estimation of Economic Relationships Using Instrumental Variables, Econometrica, 26, 393-514.

Spencer, D.E. and K.N. Berk, 1981, A Limited Information Specification Test, Econometrica, 49, 1079-1085.

White, H., 1980, Nonlinear Regression on Cross-Section Data, Econometrica, 48, 729-746.

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

White, H. and I. Domowitz, 1981, Nonlinear Regression with Dependent Observations. Manuscript, University of California at San Diego.

References

Amemiya, T., 1974, The Nonlinear Two-Stage Least-Squares Estimator, Journal of Econometrics, 2, 105-110.

Bierens, H., 1982, Consistent Model Specification Tests, Journal of Econometrics, 20.

Burguete, J.F., A.R. Gallant, and G. Souza, 1982, On Unification of the Asymptotic Theory of Nonlinear Econometric Models, Econometric Reviews, 1, 151-190.

Cox, D. R. and D. V. Hinckley, 1974, Theoretical Statistics. London: Chapman and Hall.

Chung, K. L, 1974, A Course in Probability Theory (2nd ed.). New York: Academic Press.

Domowitz, I. and H. White, 1982, Misspecified Models with Dependent Observations, Journal of Econometrics, 20, 35-58.

Ericcson, N.R., 1983, Asymptotic Properties of Instrumental Variable Statistics for Testing Non-Nested Hypotheses, Review of Economic Studies, 50, 287-304.

Gallant, A. R. and D. W. Jorgenson, 1979, Statistical Inference for a System of Simultaneous, Nonlinear, Implicit Equations in the Context of Instrumental Variables Estimation, Journal of Econometrics, 11, 275-302.

Hansen, L. P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, Econometrica, 50, 1029-1054.

Hausman, J. A., 1978, Specification Tests in Econometrics, Econometrica, 46, 1251-1272.

Hausman, J. A. and W. E. Taylor, 1980, Comparing Specification Tests and Classical Tests, Manuscript, MIT.

Hirsch, M., 1976, Differential Topology, New York: Springer-Verlag.

Holly, A., 1982a, A Remark on Hausman's Specification Test, Econometrica, 50, 749-759.

Holly, A., 1982b, A Simple procedure for Testing Whether a Subset of Endogenous Variables is Independent of the Disturbance Term in a Structural Equation, mimeo, Universite de Lausanne.