Journal of Econometrics 29 (1985) 229–256. North-Holland

GENERALIZED METHOD OF MOMENTS SPECIFICATION TESTING

Whitney K. NEWEY *

Princeton University, Princeton, NJ 08544, USA

Received July 1983, final version received March 1985

This paper analyzes the asymptotic power properties of specification tests which are based on a finite set of moment conditions. It shows that any such test may fail against general misspecification that causes estimator inconsistency. The mutual asymptotic equivalence of maximal degree of freedom tests is shown and the form of optimal tests against specific forms of misspecification is derived. Applications to testing for exogeneity of a set of instrumental variables are presented.

1. Introduction

The purpose of this paper is to analyze the asymptotic power properties of the class of specification tests which are based on a finite set of moment conditions. This class of tests is very general, since it includes both Hausman (1978) tests and Sargan (1958) and Hansen (1982) tests of overidentifying restrictions.

Section 2 lays out the framework to be used to analyze the power of specification tests. Most econometric estimators can be viewed as being obtained by minimizing a quadratic form in sample moments of functions of the data and parameters. Such an estimator will generally be consistent if each function involved, which will be referred to below as a moment function, has expectation zero at the true parameter value. Tests for violation of these moment conditions can be obtained by examining sample moments evaluated at estimated parameter values. Section 2 presents regularity conditions and obtains the asymptotic distribution of such tests under a local sequence of misspecification alternatives. The relationship of these tests to Hausman (1978) tests is also explained.

Section 3 presents results on the asymptotic power properties of moment condition tests. It is found that for any such test there are misspecification

*I would like to thank Ian Domowitz, Jim Powell, and two anonymous referees for helpful comments. Discussions with Jerry Hausman, Dan McFadden, and Ken Singleton were particularly useful. I thank the National Science Foundation for partial financial support of the work for this paper.

directions for which for some asymptotic significance level the power of the test does not go to one as the sample size grows. An equivalence result is obtained for the class of moment tests which test a full set of overidentifying restrictions. When a priori information restricts the form of misspecification, optimal moment condition tests are available. The form of these tests is presented in section 3.

Section 4 presents some applications of the general results. Testing for contamination of a subset of instrumental variables is discussed in some detail in the context of a linear equation. A Hausman (1978) test interpretation of the Hansen (1982) test of overidentifying restrictions is presented and is used to show the mutual equivalence of some tests of overidentifying restrictions in the linear simultaneous equations system.

Section 5 offers some conclusions.

2. The asymptotic distribution of GMM specification tests

Our first assumption specifies the process which generates the data.

Assumption 1. The observed data z_t, $t = 1, \ldots, T$, is $p \times 1$, consists of random vectors which are the first T elements of a strictly stationary stochastic process $\{z_t;\ t = 1, 2, \ldots\}$, and has a measurable joint density function $f(z_1, \ldots, z_T, c_T)$ with respect to a measure $\prod_{t=1}^{T} v$, where v is a σ-finite measure on R^p and c_T is a $l \times 1$ vector of parameters.

The assumption of stationarity of the data-generating process for each T rules out fixed regressors in a regression model but allows for regressors which are random draws from a fixed distribution, as might be appropriate for cross-section data. Note that the data-generating process is allowed to depend on the sample size through the parameter vector c_T, and that an extra T subscript on z_t has been suppressed for notational convenience. A problem in deriving asymptotic approximations to the distribution of specification test statistics is that if the model is misspecified the test will often reject with probability one as the sample size grows. The classical solution to this problem is to assume that correct specification occurs at $c = c_0$ and that the data-generating process is subject to Pitman (1949) drift, with $c_T = c_0 + \delta/\sqrt{T}$ for a (possibly) non-zero $l \times 1$ vector δ. This device of using a sequence of local misspecification alternatives will be the basis of most of the ensuing discussion of power properties of specification tests.

Assumption 2. $c_T = c_0 + \delta/\sqrt{T}$.

The vector of parameters c represents parameters which affect the correctness of the specification of an econometric model. For example, c might include the coefficient of an omitted variable or a vector of covariances between right-hand side (r.h.s.) variables and the disturbance. The exact meaning of the statement that the econometric model is correctly specified at c_0 will be given below as eq. (1).

Let b_0 be the $q \times 1$ vector of parameters which is to be estimated by using the data z_t, $t = 1, \dots, T$. A generalized method of moments (GMM) estimator of b_0 can be formed by using a $r \times 1$ vector of functions $g(z, b)$ of a data observation z and the parameter vector b which satisfies the following assumption. Let $f(z, c)$ be the density of a single observation.

Assumption 3. The elements of $g(z, b)$ are measurable in z for each b in a known set B, with $b_0 \in B$, and

$$\int g(z, b_0) f(z, c_0) \,\mathrm{d}v = 0. \tag{1}$$

The moment function vector $g(z, b)$ and the moment condition (1) summarize the information contained in the underlying econometric model which is used in estimation. For example, $g(z, b)$ might be a vector of cross-products of instrumental variables and disturbance terms, or it might be the gradient of the log-likelihood.

The moment condition (1) can be exploited to yield estimators of b_0 which should have good properties when the econometric model is correctly specified (i.e., when $c = c_0$). Define

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

A GMM estimator $\hat{b}_T$ of b_0 can then be obtained as the solution to

$$\min_{b \in B} g_T(b)' W_T g_T(b), \tag{2}$$

where W_T is an $r \times r$ positive semi-definite matrix which may depend on the data. The estimator $\hat{b}_T$ is obtained by setting the sample moment vector $g_T(b)$ as close as possible to zero, which is the population moment vector when $c = c_0$. This class of estimators has been considered by Amemiya (1974), Burguete, Gallant and Souza (1982), and Hansen (1982), among others.

When misspecification is present the population moment vector may not be zero, resulting in possible inconsistency of the GMM estimator $\hat{b}_T$. One method of testing for misspecification, as suggested by Hansen (1982), is to use a linear combination $L_T g_T(\hat{b}_T)$ of the estimated sample moments $g_T(\hat{b}_T)$, where L_T is a $s \times r$ matrix. When the model is correctly specified, $g_T(\hat{b}_T)$

should be close to zero in large samples by a law of large numbers and by $\hat{b}_T$ consistent for b_0. At the same time, when there are more moment functions than parameters, $g_T(\hat{b}_T)$ will not be identically zero, so that a useful test can be constructed by rejecting the null hypothesis of correct specification if a linear combination $L_T g_T(\hat{b}_T)$ is too far from zero, after approximately accounting for sampling error using asymptotic distribution theory.

To obtain the asymptotic distribution of $L_T g_T(\hat{b}_T)$ under a sequence of local alternatives, it is useful to impose regularity conditions on $g(z,b)$, the density $f(z,c)$, and to restrict the dependence across observations of the data-generating process. The set of regularity conditions which will be presented is by no means the weakest possible set of sufficient conditions, although most of the conditions should be straightforward to check in the context of a particular model.

Assumption 4. The functions $g(z,b)$ and $f(z,c)$ are continuously differentiable on B and a neighborhood C of c_0 respectively, almost everywhere v. For each $n \geq 2$ the joint density $f(z_1, z_n, c)$ is continuous in c, almost everywhere $v \times v$. Also b_0 is contained in the interior of the set B, which is compact.

The next assumption imposes dominance conditions.

Assumption 5. There exist measurable functions $a_1(z)$ and $a_2(z)$, and $d > 1$, such that almost everywhere v, and for all b in B and c in C,

$$|g(z,b)|^2 \leq a_1(z), \quad |\partial g(z,b)/\partial b| \leq a_1(z),$$

$$|\partial \ln f(z,c)/\partial c|^2 \leq a_1(z), \tag{3a}$$

$$|f(z,c)| \leq a_2(z), \quad |f(z_1,z_n,c)| \leq a_2(z_1)a_2(z_n), \quad n \geq 2, \tag{3b}$$

$$\int [a_1(z)]^d a_2(z)\,\mathrm{d}v < +\infty, \quad \int a_2(z)\,\mathrm{d}v < +\infty. \tag{4}$$

To restrict the dependence across observations of the stochastic process $\boldsymbol{Z} = \{z_t;\ t = 1,2,\dots\}$ it is useful to employ mixing conditions. Mixing conditions have recently been discussed in some detail by Domowitz and White (1982) and White and Domowitz (1984), where definitions and notation for the following assumption can be found.

Assumption 6. There exist constants $D > 0$ and λ such that, for all c in C, either (a) $\boldsymbol{Z}$ is uniform mixing with

$$\phi(m) \leq Dm^{-\lambda}, \qquad \lambda \geq d/(d-1),$$

or (b) $\boldsymbol{Z}$ is strong mixing with

$$\alpha(m) \leq Dm^{-\lambda}, \qquad \lambda \geq 2d/(d-1).$$

The next assumption is an identification condition which guarantees that the minimization problem (2) has a unique solution asymptotically. Let E denote the expectation taken at $c = c_0$, and

$$H = \mathrm{E}[\partial g(z_t, b_0)/\partial b].$$

Assumption 7. The matrix W_T has a positive semi-definite probability limit W, such that $W\mathrm{E}[g(z_t, b)] = 0$ only if $b = b_0$ and such that $H'WH$ is non-singular. Also L_T has a probability limit L with $\operatorname{rank}(L) = s$.

The asymptotic properties of the estimator $\hat{b}_T$ and the linear combination of estimated sample moments $L_T g_T(\hat{b}_T)$ are summarized in the following result. Let

$$V = \mathrm{E}\left[g(z_t, b_0) g(z_t, b_0)'\right] + \sum_{n=1}^{\infty} \left\{\mathrm{E}\left[g(z_t, b_0) g(z_{t+n}, b_0)'\right] + \mathrm{E}\left[g(z_{t+n}, b_0) g(z_t, b_0)'\right]\right\},$$

$$K = \mathrm{E}\left[g(z_t, b_0)\, \partial \ln f(z_t, c_0)/\partial c'\right],$$

$$P_W = I_r - H(H'WH)^{-1} H'W.$$

Lemma 1. If Assumptions 1–7 are satisfied, then

$$\sqrt{T}(\hat{b}_T - b_0) = -(H'WH)^{-1} H'W\sqrt{T} g_T(b_0) + \mathrm{o}_p(1), \tag{5}$$

$$\sqrt{T} L_T g_T(\hat{b}_T) = LP_W \sqrt{T} g_T(b_0) + \mathrm{o}_p(1), \tag{6}$$

$$\sqrt{T} g_T(b_0) \xrightarrow{\mathrm{d}} \mathrm{N}(K\delta, V). \tag{7}$$

Eqs. (6) and (7) imply that the asymptotic covariance matrix of $L_T g_T(\hat{b}_T)$ is given by

$$Q = LP_W V P_W' L'. \tag{8}$$

An asymptotic chi-square statistic can therefore be formed as

$$m_T = Tg_T(\hat{b}_T)'L_T'Q_T^- L_T g_T(\hat{b}_T), \tag{9}$$

where Q_T^- is an estimate of a generalized inverse [g-inverse, see Rao (1973, p. 24)] of Q. The use of a g-inverse is necessary because the singularity of P_W (note $P_W H = 0$) implies that Q may be singular.

To estimate Q, H can be estimated by

$$H_T = g_{Tb}(\hat{b}_T),$$

where $g_{Tb}(b) = \partial g_T(b)/\partial b$, P_W can be estimated by

$$P_{WT} = I_r - H_T(H_T'W_T H_T)^{-1} H_T'W_T,$$

and Q by

$$Q_T = L_T P_{WT} V_T P_{WT}' L_T',$$

where V_T is a consistent estimator of V. Detailed discussion of estimation of V is beyond the scope of this paper. The asymptotic covariance matrix estimators suggested by Hansen (1982), Domowitz and White (1982), and White and Domowitz (1984) should apply with little modification. For example, the following analog of Lemma 3.3 of Hansen (1982) holds.

Lemma 2. If Assumptions 1–7 are satisfied and $|g(z,b)|^2 \leq a_1(z)$ *for all b in* B, *then for any* $n \geq 0$

$$\frac{1}{T-n}\sum_{t=1}^{T-n} g(z_t,\hat{b}_T)g(z_{t+n},\hat{b}_T)' \xrightarrow{\mathrm{P}} \mathrm{E}\left[g(z_t,b_0)g(z_{t+n},b_0)'\right].$$

When only a finite number of terms in the asymptotic covariance matrix V of $\sqrt{T}\, g_T(b_0)$ are non-zero (e.g., cross-section data where observations may be independent), then Lemma 2 can be used to form a consistent estimator of V by replacing the non-zero terms which make up V by sample averages like those of Lemma 2. The results to be discussed later will not be restricted to this case, but will apply as long as some consistent estimator V_T of V is available.

Assumption 8. The estimator V_T satisfies $\operatorname{plim} V_T = V$, and V is non-singular.

Because of the non-uniqueness of the g-inverse, using any sequence of g-inverses of Q_T need not guarantee that Q_T^- converges in probability. One way to tie down Q_T^- is given by the following result.

Lemma 3. If Assumptions 1–8 are satisfied and S is a fixed $r \times \operatorname{rank}(Q)$ *matrix such that* $S'QS$ *is non-singular, then* $\operatorname{plim} S(S'Q_TS)^{-1}S' = Q^-$, *a g-inverse of* Q.

For example, S might be a selection matrix such that $S'QS$ is a non-singular submatrix of Q with full rank. More generally, we will make the following assumption.

Assumption 9. A sequence of g-inverses of Q_T is chosen so that $\operatorname{plim} Q_T^- = Q^-$, a g-inverse of Q.

An important special case occurs when the moment functions are linear in b, so that for a $r \times 1$ vector $g_1(z)$ and a $r \times q$ matrix $g_2(z)$ of functions

$$g(z,b) = g_1(z) + g_2(z)b. \tag{10}$$

This case will be referred to as the *linear case*.

The asymptotic distribution of the GMM specification test statistic can now be obtained.

Theorem 1. If Assumptions 1–9 are satisfied, then m_T *converges in distribution to a non-central chi-squared distribution with*

$$\operatorname{rank}(Q) = \operatorname{rank}[WH, L'] - q \tag{11}$$

degrees of freedom and non-centrality parameter

$$\lambda^2 = \delta' K' P_W' L' Q^- L P_W K \delta. \tag{12}$$

Also, if m_T *and* m_T' *are test statistics constructed with two different g-inverses satisfying Assumption 9, then* $m_T - m_T' = o_p(1)$, *while in the linear case* m_T *is numerically invariant with respect to the choice of g-inverse.*

It is possible to obtain an interesting interpretation of the non-centrality parameter λ^2. The matrix Q^- is just a g-inverse of the asymptotic covariance matrix of $L_T g(\hat{b}_T)$. The other term LP_WK which appears in eq. (12) is related to the local behavior of $L_T g(\hat{b}_T)$ under misspecification.

Theorem 2. If Assumptions 1 and 3–9 are satisfied, $g(z,b)$ *and* $f(z,c)$ *are twice continuously differentiable in* (b,c), *almost everywhere* v *with* $|\partial^2 f(z,c)/\partial c \partial c'| \le a_2(z)$, $|\partial^2 g(z,b)_i/\partial b \partial b'| \le a_1(z)$, $i = 1,\dots,r$, *and the* *data-generating process satisfies* $c_T = c$ *for some fixed* c *in a neighborhood of* c_0, *then*

$$\operatorname{plim} \hat{b}_T = b(c), \qquad \operatorname{plim} L_T g_T(\hat{b}_T) = \alpha(c),$$

such that

$$\partial b(c_0)/\partial c = -(H'WH)^{-1}H'WK, \tag{13}$$

$$\partial \alpha(c_0)/\partial c = LP_W K. \tag{14}$$

From eq. (14) we see that $LP_W K\delta$ is the directional derivative in the direction δ of the limit of $L_T g_T(\hat{b}_T)$, so that in addition to the asymptotic covariance matrix of $L_T g_T(\hat{b}_T)$ the non-centrality parameter is determined by how rapidly the limit of $L_T g_T(\hat{b}_T)$ moves away from zero as c departs from c_0.

Theorem 1 gives the asymptotic distribution under local misspecification of a very large class of specification tests. For example, tests which use one vector of moment functions for estimation and another for testing can be subsumed in this framework by choosing W_T to have certain rows and columns of zeros so that it picks out from the vector $g(z, b)$ those functions used in estimation, and by choosing L_T to pick out those used in testing. Our results also apply to Hausman (1978) specification tests.

To clarify the relationship between Hausman tests and GMM tests, which has been discussed by Ruud (1982) and White (1982), a brief discussion of Hausman tests in the GMM framework should be helpful. Let $\tilde{b}_T$ be a second GMM estimator which is obtained by solving eq. (2) with $\tilde{W}_T$ used in place of W_T, where $\tilde{W}_T$ has a limit $\tilde{W}$ which differs from W. Applying eq. (5) to both $\hat{b}_T$ and $\tilde{b}_T$ gives the asymptotic covariance matrix

$$\begin{aligned} M = {} & (H'WH)^{-1}H'WVWH(H'WH)^{-1} \\ & + (H'\tilde{W}H)^{-1}H'\tilde{W}V\tilde{W}H(H'\tilde{W}H)^{-1} \\ & - (H'WH)^{-1}H'WV\tilde{W}H(H'\tilde{W}H)^{-1} \\ & - (H'\tilde{W}H)^{-1}H'\tilde{W}VWH(H'WH)^{-1} \end{aligned} \tag{15}$$

of the difference $q_T = \hat{b}_T - \tilde{b}_T$. A consistent, positive semi-definite estimator M_T of M can be obtained by replacing H by, say H_T, W by W_T, $\tilde{W}$ by $\tilde{W}_T$, and V by V_T in eq. (15). For a g-inverse M_T^- of M_T a Hausman test statistic is then given by

$$h_T = Tq_T' M_T^- q_T. \tag{16}$$

Theorem 3. If Assumptions 1–8 are satisfied for both W_T and $\tilde{W}_T$ and a sequence of g-inverses M_T^- of M_T are chosen so that M_T^- converges in probability to M^-, a g-inverse of M, then h_T converges in distribution to a non-central chi-squared distribution with

$$\operatorname{rank}(M) = \operatorname{rank}[WH, \tilde{W}H] - q \tag{17}$$

degrees of freedom and non-centrality parameter

$$\lambda_h^2 = \delta' K' \left[WH(H'WH)^{-1} - \tilde{W}H(H'\tilde{W}H)^{-1} \right] M^- \times \left[(H'WH)^{-1} H'W - (H'\tilde{W}H)^{-1} H'\tilde{W} \right] K\delta. \tag{18}$$

Also, if h_T and h_T' are test statistics constructed with two different g-inverses satisfying the above hypothesis, then $h_T - h_T' = o_p(1)$, while in the linear case h_T is numerically invariant with respect to the choice of g-inverse.

It should be emphasized that this result gives the asymptotic distribution of most of the Hausman tests, which have been presented in the literature, when particular forms of local misspecification are present. For example, specification tests which use different moment functions, such as the test based on the difference of two weighted least-squares estimators suggested by Domowitz and White (1982), can be accommodated by stacking the functions into one vector and specifying that certain rows and columns of W_T and $\tilde{W}_T$ contain only zeros. Also, eq. (8) shows exactly how the non-centrality parameter is determined. The non-centrality parameter is a quadratic form which has a matrix M^- and, from eq. (13), a vector which is the directional derivative of the difference of the asymptotic bias of $\operatorname{plim} \hat{b}_T$ and $\operatorname{plim} \tilde{b}_T$ in the direction δ.

It is interesting to note that when $\tilde{W}$ equals V^{-1}, the asymptotic covariance matrix of $\hat{b}_T - \tilde{b}_T$ simplifies to

$$M = (H'WH)^{-1} H'WVWH(H'WH)^{-1} - (H'V^{-1}H)^{-1}, \tag{19}$$

which is the difference of asymptotic covariance matrices of $\hat{b}_T$ and $\tilde{b}_T$. As shown by Hansen (1982), choosing W equal to V^{-1} yields an estimator which is asymptotically efficient relative to any other GMM estimator, so that eq. (19) implies that for the covariance matrix M to have the simple difference form as discussed by Hausman (1978), it is sufficient that one estimator used to form $q_T = \hat{b}_T - \tilde{b}_T$ have the efficient choice of $W = V^{-1}$. For example, this observation implies that all of the specific specification tests discussed by Hausman (1978) have the simple matrix difference form even if the disturbances are not normally distributed.

The relationship between Hausman and GMM specification tests can easily be seen from a generalization to GMM estimators of the famous one-step theorems for maximum likelihood and non-linear least squares. Let

$$\bar{b}_T = \hat{b}_T - \left(H_T'\tilde{W}_T H_T\right)^{-1} H_T'\tilde{W}_T g_T(\hat{b}_T), \tag{20}$$

be an estimator obtained by starting at $\hat{b}_T$ and moving towards $\tilde{b}_T$.

Lemma 4. *If Assumptions 1–8 are satisfied for both W_T and $\tilde{W}_T$, then*

$$\sqrt{T}\left(\bar{b}_T - \tilde{b}_T\right) = o_p(1), \tag{21}$$

while $\bar{b}_T = \tilde{b}_T$ in the linear case.

This result says that the one-step estimator $\bar{b}_T$ is asymptotically equivalent to the GMM estimator $\tilde{b}_T$ so that a Hausman test based on the difference $\hat{b}_T - \tilde{b}_T$ is asymptotically equivalent to a test based on $\hat{b}_T - \bar{b}_T$. But from eq. (20),

$$\hat{b}_T - \bar{b}_T = \left(H_T'\tilde{W}_T H_T\right)^{-1} H_T'\tilde{W}_T g_T(\hat{b}_T),$$

and the term on the right-hand side is a non-singular linear combination of $H_T'\tilde{W}_T g_T(\hat{b}_T)$. Therefore a Hausman test based on the difference $\hat{b}_T - \tilde{b}_T$ is asymptotically equivalent to a GMM test with weighting matrix W and $L = H'\tilde{W}$, and numerically equivalent in the linear case. Clearly the roles of W and $\tilde{W}$ are interchangeable, so that a Hausman test based on $\hat{b}_T - \tilde{b}_T$ is also equivalent to a GMM test with weighting matrix $\tilde{W}$ and $L = H'W$.

In the next section the asymptotic distribution of GMM tests is employed to consider their asymptotic power properties.

3. Asymptotic power properties of GMM specification tests

The asymptotic power of GMM specification tests is determined by the non-centrality parameter λ^2 and the degrees of freedom, since the tail probability of a non-central chi-squared distribution is increasing in the non-centrality parameter and decreasing in the degrees of freedom. In particular, the asymptotic power curve is flat in any direction for which the non-centrality parameter is zero. In fact, as shown in Newey (1983), the set of $l \times 1$ vectors δ, such that the non-centrality parameter is equal to zero, is (under some additional regularity conditions) the tangent space at c_0 to a smooth manifold of $l \times 1$ vectors c on which the associated GMM test will not reject with probability approaching one, for some critical value.

An important property of GMM tests is that the non-centrality parameter is zero for non-zero directions δ in the presence of general forms of misspecification, so that GMM tests are not consistent against general misspecification. Bierens (1982) and Holly (1982) have both given examples of specification tests which are not consistent. For the general form of GMM test and misspecification considered in section 2, the set of directions for which the non-centrality parameter is zero can be characterized as follows.

Proposition 1. $\lambda^2 = 0$ *if and only if* $LP_W K\delta = 0$.

To interpret this result note that by V non-singular and eq. (8) the degrees of freedom of the GMM test are given by the rank of LP_W, so that the rank of $LP_W K$ is less than or equal to the degrees of freedom of the test. Noting that δ is a l-dimensional vector, it follows that the set of directions for which the non-centrality parameter is zero has a dimension greater than or equal to l minus the degrees of freedom of the test. For example, note that any particular GMM test will have a zero non-centrality parameter for some directions if the dimension of misspecification is big enough.

We can also use this result to show that when general misspecification is present there are directions for which the non-centrality parameter is zero and which result in the GMM estimator of b_0 being inconsistent. Note that

$$\partial \int g(z, b_0) f(z, c_0)\, \mathrm{d}v / \mathrm{d}c$$

$$= \int g(z, b_0)\left[\partial \ln f(z, c_0)/\partial c\right] f(z, c_0)\, \mathrm{d}v = K, \tag{22}$$

so that to say that the rank of K is equal to r means that the set of directional derivatives $K\delta$ of the expectation of the moment vector $g(z, b_0)$ traces out all of R^r, which in turn means that the misspecification allows for any direction of departure of $\mathrm{E}[g(z, b_0)]$ from zero. Also, if the rank of K is equal to r, then for any $\beta \neq 0$ we can find a δ such that $H\beta = K\delta$. For this choice of δ the non-centrality parameter is zero by $P_W H = 0$ and Proposition 1, while by eq. (13) the directional derivative of the asymptotic bias of $\hat{b}_T$ equals $-\beta \neq 0$.

The interpretation of the potential failure of GMM tests resulting from insufficient degrees of freedom is facilitated by an analogy with the Chow test in the linear model. As discussed by Rea (1978), a Chow test with insufficient degrees of freedom in the second period has a power curve which is flat on a non-trivial subspace of the alternative space. The problem there is that the parameters of the second period are not identified, so that there are non-zero values of the second-period parameters which give a test statistic which has the same distribution as if the second-period parameters were zero. The consistency problem for GMM tests is also an identification problem. The vector

$g(z, b)$ provides r moment functions (i.e., degrees of freedom) which can be used in estimating parameters. If under the alternative there are more than r parameters to estimate in the vectors b and c, then there are insufficient degrees of freedom to identify both b and c. Consequently a GMM test will not be consistent against such an alternative. In terms of Proposition 1, note that by the fact that the rank of P_W is $r-q$ the degrees of freedom of any GMM test is less than or equal to $r-q$. It follows that when r is less than $l+q$ (i.e., $r-q$ is less than l) any GMM test will have a zero non-centrality parameter for some non-zero δ.

A second important property of the class of GMM tests is the asymptotic equivalence of all such tests with $r-q$ degrees of freedom.

Proposition 2. If m_T and $\tilde{m}_T$ are two GMM test statistics with degrees of freedom $r-q$, then $m_T - \tilde{m}_T = o_p(1)$, while in the linear case $m_T = \tilde{m}_T$.

A restatement of this result is to say that a GMM test with maximal degrees of freedom $r-q$ is asymptotically equivalent to any other such GMM test. Or, if $r-q$ is identified with the number of overidentifying restrictions, a restatement is that all GMM tests which test a full set of overidentifying restrictions are mutually asymptotically equivalent.

In many situations a priori information which rules out general misspecification may be available, so that the alternative $c \neq c_0$ involves misspecification of a particular form. Important examples include contamination of particular instrumental variables, which is discussed in the next section, and violation of covariance restrictions in a simultaneous equations system [Hausman, Newey and Taylor (1983)]. In such situations it is useful to have available GMM tests which are optimal against a particular form of misspecification. The notion of optimality which we use here is similar in nature to the idea that a GMM estimator with weighting matrix $W = V^{-1}$ is optimal in the class of GMM estimators. A GMM test will be referred to as optimal in the class of GMM tests if it has the largest possible value of the non-centrality parameter for all δ in R^l and has the smallest possible degrees of freedom among tests with this property.

One form of optimal GMM test can be formed using the GMM estimator $\hat{b}_T$ with $W = V^{-1}$. Suppose for the moment that there is available an estimator K_T of K which is consistent when $c = c_0$. Consider a GMM test with the $l \times r$ linear combination matrix $L_T = K_T' V_T^{-1}$. Straightforward calculation using eq. (8) shows that the asymptotic covariance matrix of $L_T g_T(\hat{b}_T)$ is given by

$$Q = K'V^{-1}K - K'V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1}K. \tag{23}$$

Let Q_T be obtained by replacing H, V and K by $g_{Tb}(\hat{b}_T)$, V_T and K_T, respectively, and let $m_T = g_T(\hat{b}_T)'L_T'Q_T^{-1}L_T g_T(\hat{b}_T)$ be the corresponding GMM statistic. Under the conditions of the following theorem Q will be non-singular and m_T will give an optimal GMM test.

Proposition 3. If $\operatorname{plim} K_T = K$ *and* $[H, K]$ *has* $\operatorname{rank} q + l$, *then* m_T *has* l *degrees of freedom and is an optimal GMM test statistic.*

The form of this test statistic has a straightforward interpretation. If $g_T(b)$ is thought of as a vector of residuals, then the estimator $\hat{b}_T$ can be interpreted as a generalized (non-linear) least squares (GLS) estimator. Then the linear combination of residuals $K_T' V_T^{-1} g_T(\hat{b}_T)$ is the estimated score vector which would be used to form a score test for the inclusion of the variables K_T if the disturbance vector $g_T(b_0)$ had a normal distribution. This procedure is asymptotically optimal because $\sqrt{T}\, g_T(b_0)$ converges in distribution to a normal random vector. Note that, in general, consistent estimation of K may be difficult since it can require knowledge of the form of the density $f(z, c)$. In the examples considered below this problem does not arise.

An alternative form of an optimal GMM test can be obtained using an estimator which partials out K_T. Let

$$\tilde{W} = V^{-1} - V^{-1}K(K'V^{-1}K)^{-1}K'V^{-1}, \tag{24}$$

let $\tilde{W}_T$ be obtained from $\tilde{W}$ by replacing K by K_T and V by V_T, and let $\tilde{b}_T$ be the GMM estimator with weighting matrix $\tilde{W}_T$. This estimator has a straightforward interpretation if $g_T(b)$ is thought of as a vector of residuals. The estimator is then a GLS estimator when K_T is included in the residuals as a $r \times l$ matrix of r observations on l variables. In terms of the local results of Theorem 2 it can be shown that $\tilde{b}_T$ is an optimal GMM estimator among those GMM estimators which have a zero derivative of the asymptotic bias at $c = c_0$, i.e., those estimators for which eq. (13) equals zero.

Straightforward calculation using eq. (8) can be used to show that the asymptotic covariance matrix of $L_T g_T(\tilde{b}_T)$ is given by

$$\tilde{Q} = K'V^{-1}K + K'V^{-1}H(H'\tilde{W}H)^{-1}H'V^{-1}K, \tag{25}$$

where we again let $L_T = K_T' V_T^{-1}$. Let $\tilde{Q}_T$ be obtained from $\tilde{Q}$ by replacing H, V and K by $g_{Tb}(\tilde{b}_T)$, V_T and K_T, respectively, and let $\tilde{m}_T = g_T(\tilde{b}_T)' L_T' \tilde{Q}_T^{-1} L_T g_T(\tilde{b}_T)$ be the associated GMM statistic with $L_T = K_T' V_T^{-1}$. Under the conditions of the following theorem $\tilde{Q}$ will be non-singular and $\tilde{m}_T$ will give an optimal GMM test statistic which is asymptotically equivalent to the previously presented optimal GMM test.

Proposition 4. If $\operatorname{plim} K_T = K$ *and* $[H, K]$ *has rank* $q + l$, *then* $\tilde{m}_T$ *has* l *degrees of freedom and* $m_T - \tilde{m}_T = o_p(1)$, *while in the linear case* $m_T = \tilde{m}_T$.

It is also possible to form a Hausman test statistic based on the estimator difference $q_T = \tilde{b}_T - \hat{b}_T$. Using eq. (19) it follows that the asymptotic covariance matrix of q_T is equal to

$$M = (H'\tilde{W}H)^{-1} - (H'V^{-1}H)^{-1}. \tag{26}$$

Let M_T be obtained from M by replacing V and K by V_T and K_T, respectively, and H by, say $g_{Tb}(\tilde{b}_T)$, and let $h_T = Tq_T'M_T^- q_t$ be the associated Hausman statistic. The use of a generalized inverse is called for because M may be singular. The analysis of Holly (1982a) and Hausman and Taylor (1980) can be generalized to give the following result.

Proposition 5. If plim $K_T = K$ *and* $[H, K]$ *has rank* $q + l$, *then* h_T *has degrees of freedom equal to the rank of* $K'V^{-1}H$. *Also, if* rank$[K'V^{-1}H] = l$, *then* $\tilde{m}_T - h_T = o_p(1)$, *with numerical equality in the linear case, while if* rank $[K'V^{-1}H] < l$, *then the asymptotic power curves of* $\tilde{m}_T$ *and* h_T *cross.*

It is useful to compare the optimal GMM tests presented above with the class of asymptotically equivalent $r - q$ degrees of freedom tests which test a full set of overidentifying restrictions.

Proposition 6. For all δ *in* R^l *the non-centrality parameter for a GMM test with* $r - q$ *degrees of freedom is greater than or equal to the non-centrality parameter for any other GMM test.*

It follows that the non-centrality parameter for the $r - q$ degrees of freedom tests is equal to that of the optimal tests given above. These optimal tests will have larger local power than the $r - q$ degrees of freedom test when the degrees of freedom l of the optimal tests is less than $r - q$, i.e., when the dimension of misspecification is less than the number of overidentifying restrictions.

One important situation where optimal GMM tests are straightforward to construct and should be useful is when misspecification results in contamination of a subset of moment functions. For example, such a situation can arise when certain instrumental variables are suspect. Eichenbaum, Hansen and Singleton (1984) also discuss tests of a subset of moment functions and present empirical applications which involve testing the validity of instrumental variables in a rational expectations model.

Suppose that the $r \times 1$ vector of moment functions is partitioned as $g(z, b) = (g_1(z, b)', g_2(z, b)')'$, where $g_2(z, b)$ is an $l \times 1$ vector and

$$\int g(z, b_0) f(z, c)\, \mathrm{d}v = (0, c')'. \tag{27}$$

Thus we parameterize the misspecification as the expectation of $g_2(z, b_0)$ and under misspecification the moment condition $\mathrm{E}[g_1(z_t, b_0)] = 0$ remains satisfied. Then from eq. (22) it follows that $K = [0, I_l]'$, where I_l is a l-dimensional identity matrix. Here the problem of obtaining a consistent estimator of K does not arise and optimal GMM tests can be calculated without further knowledge of the data generating process.

To see what form optimal GMM tests of a subset of moment functions take, partition H, V and V^{-1} conformably with $g(z, b)$,

$$H = [H_1', H_2']', \quad V = [V_{ij}], \quad V^{-1} = [V^{ij}], \quad i, j = 1,2. \tag{28}$$

Straightforward calculation shows that the linear combination matrix $L = K'V^{-1}$ becomes

$$L = [0, I_l]V^{-1} = V^{22}[-V_{21}V_{11}^{-1}, I_l]. \tag{29}$$

Since V^{22} is non-singular it can be dropped without affecting the test and an equivalent choice of linear combination matrix is

$$\tilde{L}_T = [-V_{T21}V_{T11}^{-1}, I_l], \tag{30}$$

where V_T is also partitioned conformably with $g(z, b)$. With this choice the linear combination of estimated moment functions used in the optimal GMM test based on $\hat{b}_T$ is equal to $g_{T2}(\hat{b}_T) - V_{T21}V_{T11}^{-1}g_{T1}(\hat{b}_T)$, so that the optimal linear combination matrix partials out the uncontaminated moment functions.

Further straightforward calculation shows that the weighting matrix $\tilde{W}$ for the estimator $\tilde{b}_T$ is a block-diagonal matrix, $\tilde{W} = \mathrm{diag}[V_{11}^{-1}, 0]$. Thus, $\tilde{b}_T$ is the optimal GMM estimator among those GMM estimators which use only the moment function vector $g_1(z, b)$. Such estimators will remain consistent for b_0 under misspecification when misspecification does not affect the moment condition $\mathrm{E}[g_1(z_t, b_0)] = 0$. Note that the hypothesis that the rank of $[H, K]$ is $q + l$ of Propositions 4–6 here requires that $\mathrm{rank}(H_1) = q$, which is an identification assumption for b with the subvector $g_1(z, b_0)$ of moment functions. The asymptotic covariance matrices of $\tilde{L}_T g_T(\hat{b}_T)$ and $\tilde{L}_T g_T(\tilde{b}_T)$ respectively are

$$Q = V_{22} - V_{21}V_{11}^{-1}V_{12} - B(H'V^{-1}H)^{-1}B', \tag{31a}$$

$$\tilde{Q} = V_{22} - V_{21}V_{11}^{-1}V_{12} + B(H_1'V_{11}^{-1}H_1)^{-1}B', \tag{31b}$$

where $B = H_2 - V_{21}V_{11}^{-1}H_1$. As before consistent estimators of these asymptotic covariance matrices can be obtained using consistent estimators of V and H.

The other test which we have presented as a test for a specific form of misspecification is a Hausman test based on the difference of the GMM estimators $\tilde{b}_T$ and $\hat{b}_T$. The asymptotic covariance matrix of $q_T = \tilde{b}_T - \hat{b}_T$ is

$$M = \left(H_1'V_{11}^{-1}H_1\right)^{-1} - \left(H'V^{-1}H\right)^{-1}. \tag{32}$$

A consistent estimator of M can be obtained by replacing V by V_T and H by, say $g_{Tb}(\tilde{b}_T)$. Unlike Q and $\tilde{Q}$, the matrix M may be singular, so that use of a generalized inverse may be called for.

4. Applications

For a first application of the theoretical results of sections 2 and 3 consider a GMM test statistic $\overline{m}_T = Tg_T(\hat{b}_T)'V_T^{-1}g_T(\hat{b}_T)$, where $\hat{b}_T$ is the optimal GMM estimator with $W = V^{-1}$. It is straightforward to check that the degrees of freedom of this test is $r-q$ and that V^{-1} is a g-inverse of the asymptotic covariance matrix of $g_T(\hat{b}_T)$. This test statistic was suggested by Hansen (1982) as a convenient test for the overidentifying restrictions embodied in the moment condition $\mathrm{E}[g(z_t, b_0)] = 0$. The flat asymptotic power curve of GMM tests in certain directions (Proposition 1) indicates that this statistic fails, along with every other GMM test (including Hausman tests), to be an omnibus test for misspecification. Nevertheless $\overline{m}_T$, or any other GMM test statistic with $r-q$ degrees of freedom, comes closest in the class of GMM tests to being an omnibus test. For any direction of misspecification the non-centrality parameter is as large as possible (Proposition 6).

It is also interesting that in certain circumstances $\overline{m}_T$ can be interpreted as a Hausman test, as conjectured by Mankiw, Rotemberg and Summers (1982). When $r-q$ exceeds q the degrees of freedom of $\overline{m}_T$ will differ from that of any Hausman test based on the difference of two estimators of b_0, but otherwise the equivalence of Hausman and GMM tests (Lemma 4) and the mutual asymptotic equivalence of all GMM tests with $r-q$ degrees of freedom (Proposition 2) imply that $\overline{m}_T$ will be asymptotically equivalent to *any* Hausman test with $r-q$ degrees of freedom and numerically equivalent in the linear case. For example, in the context of a linear simultaneous equations system estimated by instrumental variables, $\overline{m}_T$ equals T times the Gallant and Jorgenson (1979, p. 279) testing criteria, evaluated at the three-stage least-squares (3SLS) estimates. If the Hausman (1978) test based on the difference of the two-stage least-squares (2SLS) and 3SLS estimators has the same degrees of freedom it will equal $\overline{m}_T$, when the same disturbance covariance matrix estimator is used throughout.

To consider an example in some detail, let a linear equation be given by

$$y_t = Z_t b_0 + u_t, \qquad t = 1, \dots, T, \tag{33}$$

where Z_t is a $1 \times q$ vector and u_t a disturbance term. Let a $1 \times r$ vector of instrumental variables be given by X_t and suppose that if the equation is correctly specified the orthogonality condition $\mathrm{E}[X_t' u_t] = 0$ is satisfied. The moment condition vector for this problem is then given by

$$g(z_t, b) = X_t'(y_t - Z_t b),$$

where $z_t = (y_t, Z_t, X_t)$. For simplicity assume that observations are independently distributed and that there is no heteroskedasticity. Let $\sigma^2 = \mathrm{E}[u_t^2 | X_t]$. Note that in terms of the notation of sections 2 and 3, $H = -\mathrm{E}(X_t' Z_t)$ and $V = \sigma^2 \mathrm{E}(X_t' X_t)$. To guarantee that the regularity conditions of section 2 are satisfied assume that $\mathrm{E}(X_t' X_t)$ is non-singular, that $\mathrm{rank}[\mathrm{E}(X_t' Z_t)] = q$, and that $\int |z|^{2d} a_2(z)\,\mathrm{d}v$ is finite for some $d > 1$.

For this example the optimal GMM estimator is the 2SLS estimator

$$\hat{b} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'y, \tag{34}$$

where $y = (y_1, \ldots, y_T)'$, $Z = (Z_1', \ldots, Z_T')'$, $X = (X_1', \ldots, X_T')'$, $\hat{N} = X(X'X)^{-1}X'$, and $\hat{Z} = \hat{N}Z$. For notational simplicity the T subscript on $\hat{b}$ is dropped. Consistent estimators of V and H are given by $V_T = \sigma^2 X'X/T$ and $H_T = -X'Z/T$, respectively, where $\hat{\sigma}^2 = \hat{u}'\hat{u}/(T-q)$ and $\hat{u} = y - Z\hat{b}$. The previous results on numerical equality of various test statistics depend on using the same estimator of V to form each test statistic. Consequently the numerical equality results given below will not hold when different estimators of σ^2 are used, although the relevant statistics will remain asymptotically equivalent as long as each estimator of σ^2 is consistent.

In this context, where $g_T(\hat{b}_T) = X'\hat{u}/T$, the Hansen (1982) statistic is

$$\overline{m}_T = T g_T(\hat{b}_T)' V_T^{-1} g_T(\hat{b}_T) = (T-q)\hat{u}'\hat{N}\hat{u}/\hat{u}'\hat{u}.$$

As noted by Hausman (1984), $\overline{m}_T = (T-q)R^2$, where R^2 is the uncentered r-squared from a regression of $\hat{u}$ on X. There are a surprisingly large class of statistics which are equal to $\overline{m}_T$. Any GMM or Hausman test statistic with $r - q$ degrees of freedom is numerically equal to $\overline{m}_T$ (Lemma 4 and Proposition 2). For example, consider the test proposed by Hausman and Taylor (1980) based on the difference of $\hat{b}$ and a 2SLS estimator

$$\tilde{b} = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'y, \tag{35}$$

which uses only a subset of the instrumental variables, where $X_t = (X_{t1}, X_{t2})$, $\mathrm{E}(X_{t1}' Z_t)$ has rank q, X_{t2} is a $1 \times l$ subvector of X_t, $X = [X_1, X_2]$ is partitioned conformably with X_t, $\tilde{N} = X_1(X_1'X_1)^{-1}X_1'$, and $\tilde{Z} = \tilde{N}Z$. A Hausman test statistic based on the difference of $\hat{b}$ and $\tilde{b}$ is given by

$$h_T = (\tilde{b} - \hat{b})'\left[(\tilde{Z}'\tilde{Z})^{-1} - (\hat{Z}'\hat{Z})^{-1}\right]^{-}(\tilde{b} - \hat{b})/\hat{\sigma}^2.$$

Note that h_T is invariant with respect to g-inverse by Theorem 3. As shown in Hausman and Taylor (1980) h_T has $\min\{l, q-s\}$ degrees of freedom, where s is the number of right-hand side (r.h.s.) variables (components of Z_t) which are retained as instrumental variables (components of X_{t1}) when $\tilde{b}$ is formed. When $r-q=\min\{l, q-s\}$ it follows that $h_T=\overline{m}_T$. Remarkably, this equality holds *independently* of the particular instrumental variables which are excluded when forming $\tilde{b}$.

When a priori information restricts misspecification to contamination of a particular subset of instrumental variables, the discussion in section 3 can be used to obtain optimal GMM tests. Suppose that misspecification takes a form such that when the model is misspecified

$$\mathrm{E}(X_{t1}'u_t)=0, \qquad \mathrm{E}(X_{t2}'u_t)=c, \tag{36}$$

with a corresponding partition of the vector of moment functions $g_1(z_t, b)=X_{t1}'(y_t-Z_tb)$ and $g_2(z_t, b)=X_{t2}'(y_t-Z_tb)$. The optimal linear combination matrix is $\tilde{L}_T=[-X_2'X_1(X_1'X_1)^{-1}, I_l]$, so that the optimal GMM test based on $\hat{b}$ will have a linear combination of estimated moment functions

$$T\tilde{L}_Tg_T(\hat{b})=X_2'\hat{u}-X_2'X_1(X_1'X_1)^{-1}X_1'\hat{u}=X_2'(I-\tilde{N})\hat{u}=R'\hat{u},$$

where $R=(I-\tilde{N})X_2$ is the $T\times l$ matrix of residuals from a regression of the columns of the contaminated variables X_2 on the uncontaminated variables X_1. Also,

$$T\left[H_{T2}-V_{T12}V_{T11}^{-1}H_{T1}\right]=-R'Z=-R'\hat{Z},$$

where the second equality holds because $\hat{N}R=R$, and

$$T\left[V_{T22}-V_{T12}V_{T11}^{-1}V_{T12}\right]=\hat{\sigma}^2R'R, \qquad TH_T'V_T^{-1}H_T=\hat{Z}'\hat{Z}/\hat{\sigma}^2.$$

Then from the formula for Q in eq. (31a) it follows that a GMM test statistic is given by

$$m_T=\hat{u}'R\left[R'R-R'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'R\right]^{-1}R'\hat{u}/\hat{\sigma}^2. \tag{37}$$

By Proposition 3 this statistic is an optimal GMM statistic for misspecification of the form in eq. (36) and has l degrees of freedom.

It is useful to note that m_T can be computed via a regression. By the fact that $\hat{Z}'(y-\hat{Z}\hat{b})=\hat{Z}'\hat{N}(y-\hat{Z}\hat{b})=\hat{Z}'\hat{u}=0$ and $R'(y-\hat{Z}\hat{b})=R'\hat{N}(y-\hat{Z}\hat{b})=R'\hat{u}$ it follows that the statistic m_T is the score statistic for the hypothesis $\delta=0$ in the regression equation

$$y=\hat{Z}b+R\delta+w, \tag{38}$$

with the important modification that the estimate of the variance of the disturbance term in this equation is replaced by $\hat{\sigma}^2$. By the usual numerical equivalence of the score test with the Wald and F tests when the same estimate of the disturbance variance is used throughout [e.g., Engle (1984)], m_T can also be obtained using these other statistics while replacing the estimate of the disturbance variance in eq. (38) by $\hat{\sigma}^2$.

The Hausman statistic h_T can also be obtained via a regression. By Lemma 4 h_T is equal to a GMM statistic with $W_T = V_T^{-1}$ and $L_T = H_T' \cdot \operatorname{diag}[V_{T11}^{-1}, 0]$. This GMM statistic has a linear combination of estimated moment functions given by

$$TL_T g_T(\hat{b}) = \left[Z'X_1(X_1'X_1)^{-1}, 0\right] X'\hat{u} = Z'\tilde{N}\hat{u} = \tilde{Z}'\hat{u}.$$

Then using eq. (8) to form a consistent estimator of the asymptotic covariance matrix of $TL_T g_T(\hat{b})$ it follows that

$$\begin{aligned} h_T &= \hat{u}'\tilde{Z}\left[\tilde{Z}'\tilde{Z} - \tilde{Z}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\tilde{Z}\right]^{-}\tilde{Z}'\hat{u}/\hat{\sigma}^2 \\ &= \hat{u}'\tilde{Z}S\left[S'\tilde{Z}'\tilde{Z}S - S\tilde{Z}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\tilde{Z}S\right]^{-1}S'\tilde{Z}'\hat{u}/\hat{\sigma}^2, \end{aligned} \tag{39}$$

where S is chosen so that $\operatorname{rank}[\hat{Z}, \tilde{Z}S] = q + \min\{q - s, l\}$. The second equality is obtained by choosing a particular g-inverse, as discussed in section 2 (see Lemma 3). Similarly to the discussion of eq. (38) it follows that h_T can be computed by obtaining the score, Wald, or F test for the hypothesis $\tilde{\delta} = 0$ in the regression equation

$$y = \hat{Z}b + \tilde{Z}S\tilde{\delta} + \bar{w}, \tag{40}$$

and replacing the estimate of the disturbance variance by $\hat{\sigma}^2$.

It appears that the optimal GMM test statistic m_T for the validity of a subset of instrumental variables is new, although it is related to previously proposed tests in special cases. When $l = r - q$, m_T is equal to the Hansen (1982) statistic. When the contaminated instrumental variables are r.h.s. variables (i.e., the columns of Z contain X_2) the statistic m_T will test whether or not some r.h.s. variables are correlated with the disturbance term, while allowing other r.h.s. variables to be endogenous under the null hypothesis. Tests of such a null hypothesis have been specifically considered by Spencer and Berk (1981), Holly (1982b), and others [see Holly (1982b) for references]. To see the relationship of m_T to these tests, note that h_T is a Hausman test based on the difference of the optimal GMM estimator which uses all the moment functions (i.e., $\hat{b}$) and the optimal GMM estimator which uses only the uncomtaminated moment functions (i.e., $\tilde{b}$). Then when Z includes X_2, the number of variables in X_2 plus the number of variables common to Z and X_1 is no greater than q (i.e., $l+s \leq q$), so that the degrees of freedom of h_T is l, and it follows from Proposition 5 that $m_T = h_T$. Then since h_T is the Hausman test discussed by Spencer and Berk (1981) and Holly (1982b) m_T is equal to these tests. It is interesting to note that the regression equation, eq. (38), is different from the expanded regression of Holly (1982b) and that the discussion in Spencer and Berk (1981) of a regression method for computing h_T is incomplete. With respect to Spencer and Berk (1981), it follows that in their equation (13) only l extra variables should be tested for significance and the usual estimates of the disturbance variance for this equation should be replaced by an estimator of the disturbance of the structural equation [their equation (3)].

It is straightforward to obtain optimal GMM tests for the validity of certain instrumental variables in the presence heteroskedasticity and/or autocorrelation. To obtain these tests we simply replace $\hat{\sigma}^2 X'X/T$ by an estimator V_T of the asymptotic covariance matrix of $X'u/\sqrt{T}$ which is appropriate in the presence of heteroskedasticity and/or autocorrelation [see White and Domowitz (1984) for such a choice of V_T]. The optimal GMM estimator of b_0 in eq. (33) is then

$$\hat{b} = \left(Z'XV_T^{-1}X'Z\right)^{-1} Z'XV_T^{-1}X'y.$$

With a choice of V_T which is appropriate for the heteroskedastic case this estimator is White's (1981) two-stage instrumental-variables estimator, while in the autocorrelation case it is the estimator of Cumby, Huizinga and Obstfeld (1983). The general formulae of section 3 then apply, with $g_T(b) = X'(y - Zb)/T$ and $H_T = -X'Z/T$, to obtaining optimal GMM test statistics for misspecification of the form given in eq. (36).

5. Conclusion

The fact that moment condition tests are not consistent against general misspecification indicates that some caution may be justified when interpreting the results of such tests. If the test does not result in rejection of the null hypothesis of no misspecification, it may be because the test has low power against a particular alternative, even though this alternative causes parameter inconsistency. This inconsistency result may also be a reason to pursue the work begun by Bierens (1982) in order to have available simple, omnibus misspecification tests for situations in which little information is available on the form of misspecification.

The asymptotic equivalence of all moment condition tests with maximal degrees of freedom is convenient, because it allows us to limit some of the discussion of specification tests. Due to the fact that all moment tests with maximal degrees of freedom are asymptotically equivalent, we can perhaps limit our use of such tests to the most convenient member of this group of tests.

Optimal moment condition tests should prove to be useful in cases where a priori information restricts the form of misspecification. Optimal tests for the violation of specific moment conditions can be used to formulate tests for the validity of instrumental variables, as considered in section 4, and tests of restrictions on the disturbance matrix in a linear simultaneous equations system, as considered by Hausman, Newey and Taylor (1983). These optimal tests are more closely related to tests of parametric hypotheses (e.g., the Wald test) than other specification tests, precisely because optimal tests are formulated to have good power against certain alternatives.

Appendix

We first give several lemmas which are useful in the proofs that follow. Let $R(B)$ be the rank of a matrix B.

Lemma A.1. [*Rao (1973, 1.b.5,(iv),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 A.2. [*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 A.3. [*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 A.4. For conformable matrices A and B, $A(A'A)^{-}A'$ and $A(A'A)^{-}A' - AB(B'A'AB)^{-}B'A$ are idempotent for any g-inverse choices.

Lemma A.5. Let A be a $k \times l$ matrix, B a $l \times m$ matrix, and C a $l \times n$ matrix. If the columns of C form a basis for the column nullspace of A and $R(B) = m$, then $R(AB) = R([C, B]) - n$.

Lemma A.6. For conformable matrices A and B, if B is positive definite, then $R(A'(ABA')^{-}A) = R(A)$ for any choice of g-inverse.

Let $\{c_T\}_{T=1}^{\infty}$ be a sequence contained in C which converges to $\bar{c}$ in C and let E_T and cov_T denote the expectation and covariance, respectively, taken at $f(z, c_T)$, and let $\bar{\mathrm{E}}$ denote the expectation taken at $f(z, \bar{c})$. Unless noted otherwise lim(·) and plim(·) will denote the limit and probability limit, respectively, as $T \to \infty$.

Lemma A.7. If Assumptions 1, 5, and 6 are satisfied and $w(z,b)$ is a function which is measureable in z for all b, continuous on B almost everywhere v, and $\sup_B|w(z,b)| \le a_1(z)$, then for $w_t(b) = w(z_t, b)$, $\mathrm{E}_T[w_1(b)]$ is continuous on B uniformly in T and converges to $\bar{\mathrm{E}}[w_1(b)]$ uniformly on B and

$$\operatorname{plim} \sup_B \left| \sum_{t=1}^{T} w_t(b)/T - \bar{\mathrm{E}}[w_1(b)] \right| = 0. \tag{A.1}$$

The proofs of Lemmas A.1–A.7 are omitted for brevity, but are available upon request from the author.

Let

$$h(b,c) = \int g(z,b) f(z,c)\,\mathrm{d}v.$$

Lemma A.8. If Assumptions 1–6 are satisfied, then

$$\sqrt{T}\,[g_T(b_0) - h(b_0, c_T)] \xrightarrow{\mathrm{d}} \mathrm{N}(0, V).$$

Proof. Let $w_{tT} = \lambda'[g(z_t, b_0) - h(b_0, c_T)]$, where z_t is the tth observation of the stochastic process with $c = c_T$ and λ is any $r \times 1$ vector with $\lambda'\lambda = 1$. Then for each T

$$\mathrm{E}_T(w_{tT}) = 0, \qquad t = 1,2 \ldots. \tag{A.2}$$

Next, stationarity implies that for each T and non-negative integer a,

$$N_a(T) = \mathrm{E}_T\left[\left(\sum_{t=1+a}^{T+a} w_{tT}/\sqrt{T}\right)^2\right] = \alpha_{1T} + 2\sum_{n=2}^{\infty} \alpha_{nT}, \tag{A.3}$$

where $\alpha_{nT} = \max\{0, (T-n+1)/T\}\operatorname{cov}_T(w_{1T}, w_{nT})$, $n \ge 1$. By Lemma A.7, $h(b_0, c_T)$ converges to $h(b_0, c_0) = 0$. Also, by Assumptions 4, 5 and 6 and the dominated convergence theorem (DCT) $\mathrm{E}_T[g(z_1, b_0) g(z_n, b_0)']$ converges to

$E[g(z_1, b_0)g(z_n, b_0)']$. It follows that

$$\lim_{T\to\infty} \alpha_{nT} = \lambda' E\left[g(z_1, b_0)g(z_n, b_0)'\right]\lambda, \qquad n \geq 1. \tag{A.4}$$

From Assumptions 5 and 6 and Lemma 2.2 of White and Domowitz (1984) there are finite, constant D', γ, with $\gamma > 1$, such that for all positive integers T and n, $|\alpha_{nT}| \leq D'n^{-\gamma}$. Then by eq. (A.4) and the DCT applied to the counting measure on the integers $\lim_{T\to\infty} N_a(T) = \lambda' V \lambda$, where the convergence is uniform in a because $N_a(T)$ does not depend on a.

Next, note that by Assumption 5 there exists a finite constant D'', such that for all T and t

$$E_T|w_{tT}|^{2d} \leq D''. \tag{A.5}$$

Finally, Assumption 6 implies that all the hypotheses of Theorem 2.4 of White and Domowitz (1984) are satisfied uniformly in T. Examination of the proof of this theorem, which uses inequalities based on mixing and moment conditions, leads to the conclusion that, even though the stochastic process w_{tT} depends on T,

$$\sum_{t=1}^{T} w_{tT}/\sqrt{T} \xrightarrow{d} N(0, \lambda' V \lambda), \tag{A.6}$$

so that the conclusion follows by the Cramer–Wold device.

Proof of Lemma 1. To prove eq. (7), Assumptions 4 and 5 imply that $h(b_0, c)$ is continuously differentiable in c and that differentiation and integration can be interchanged. A mean value expansion gives

$$\begin{aligned} \lim\left[\sqrt{T} h(b_0, c_T)\right] &= \lim\left[\sqrt{T} h(b_0, c_0) + \partial h(b_0, \tilde{c}_T)/\partial c \cdot \sqrt{T}(c_T - c_0)\right] \\ &= \partial h(b_0, c_0)/\partial c \cdot \delta = K\delta, \end{aligned} \tag{A.7}$$

where $\tilde{c}_T$ lies on the line joining c_T and c_0 so that $\lim(\tilde{c}_T) = c_0$. Eq. (7) now follows from Lemma A.8.

To show that eq. (5) holds, note that by Assumption 5 and Lemma A.7, $g_T(b)$ converges in probability to $h(b, c_0)$ uniformly in b, so that $g_T(b)'W_T g_T(b)$ converges uniformly to $h(b, c_0)'Wh(b, c_0)$. By Assumption 7, $h(b, c_0)'Wh(b, c_0)$ has a unique minimum at b_0 in B. Therefore a convergence in probability version of Lemma 3 of Amemiya (1973) implies $\operatorname{plim} \hat{b}_T = b_0$. Then by Assumption 4, the first-order condition

$$g_{Tb}(\hat{b}_T)' W_T \sqrt{T} g_T(\hat{b}_T) = 0, \tag{A.8}$$

will be satisfied with probability approaching one. Expanding $\hat{b}_T$ around b_0,

$$\sqrt{T}\,g_T(\hat{b}_T)=\sqrt{T}\,g_T(b_0)+g_{Tb}(\dot{b}_T)\sqrt{T}(\hat{b}_T-b_0), \tag{A.9}$$

where $\dot{b}_T$ lies on the line joining $\hat{b}_T$ and b_0, with $\operatorname{plim}\dot{b}_T=b_0$. By Lemma A.7, Assumption 5, and Lemma 4 of Amemiya (1973) it follows that

$$\operatorname{plim} g_{Tb}(\dot{b}_T)=\operatorname{plim} g_{Tb}(\hat{b}_T)=H. \tag{A.10}$$

Eq. (5) now follows from eqs. (A.8)–(A.10), $\operatorname{plim}W_T=W$, $H'WH$ non-singular, and $\sqrt{T}\,g_T(b_0)$ bounded in probability. Eq. (6) then follows from eqs. (A.9), (A.10), and (5).

Proof of Lemma 2. This result follows by Assumption 6, Lemma A.7 with $w_t(b)=g(z_t,b)g(z_{t+n},b)$, and Lemma 4 of Amemiya (1973).

Proof of Lemma 3. $Q^-=S(S'QS)^{-1}S'$ follows from Lemma A.2. Then $\operatorname{plim}Q_T^-=Q^-$ follows by $\operatorname{plim}Q_T=Q$ and continuity of matrix inversion.

Proof of Theorem 1. By eq. (6) and Assumption 8,

$$m_T=Y_T'P_W'L'Q^-LP_WY_T+o_p(1), \tag{A.11}$$

where $Y_T=\sqrt{T}\,g_T(b_0)$. The limiting non-central chi-square distribution of m_T, with $R(Q)$ degrees of freedom, then follows by eqs. (6) and (7). Then because WH is a basis for the nullspace of P_W' and $R(Q)=R(P_W'L')$, eq. (11) follows from Lemma A.5. Asymptotic equivalence of m_T and m_T' is implied by eq. (A.11) and Lemma A.3. In the linear case let $\bar{g}_1=\sum_{t=1}^T g_1(z_t)/T$ and $\bar{g}_2=\sum_{t=1}^T g_2(z_t)/T$. Then for $H_T=\bar{g}_2$ it follows that

$$m_T=T\bar{g}_1'P_{WT}'L_T'\left(L_TP_{WT}V_TP_{WT}'L_T'\right)^-L_TP_{WT}\bar{g}_1. \tag{A.12}$$

Invariance of m_T with respect to g-inverse in the linear case now follows from Lemma A.3.

Proof of Theorem 2. By the hypotheses of this theorem and Assumption 5, $h(b,c)$ is twice continuously differentiable in (b,c). Since $h(b_0,c_0)=0$ and $\partial h(b_0,c_0)/\partial b=H$, the Hessian matrix of $\Psi(b,c)=h(b,c)'Wh(b,c)/2$ in b at (b_0,c_0) is $H'WH$, which is non-singular. Define $b(c)$ as the value of b in B which minimizes $\Psi(b,c)$. By the implicit function theorem applied to the first-order condition for this problem, and by compactness of B and continuity of $h(b,c)$, it follows that $b(c)$ is unique for c in a small enough neighborhood of c_0 and $b(c)$ satisfies eq. (13). The conclusion then follows by Lemma A.7

applied to $g_T(b)$ with c fixed in this neighborhood, by Lemma 3 of Amemiya (1973), and the chain rule applied to $\alpha(c) = \operatorname{plim} g_T(\hat{b}_T) = h(b(c), c)$.

Proof of Theorem 3. Let $D = (H'WH)^{-1}H'W - (H'\tilde{W}H)^{-1}H'\tilde{W}$, and note that $M = DVD'$. By eq. (5) and $\operatorname{plim} M_T^- = M^-$ it follows that

$$h_T = Y_T'D'M^-DY_T + o_p(1). \tag{A.13}$$

The conclusion now follows as in the proof of Theorem 1, noting that in the linear case, where D_T is obtained from D by replacing H, W and $\tilde{W}$ by H_T, W_T and $\tilde{W}_T$, respectively, h_T is given by

$$h_T = T\bar{g}_1'D_T'\left(D_TV_TD_T'\right)^-D_T\bar{g}_1. \tag{A.14}$$

Proof of Lemma 4. By the definition of $\bar{b}_T$, eq. (A.9), and eq. (5),

$$\begin{aligned}\sqrt{T}(\bar{b}_T - \tilde{b}_T) &= \left[I - \left(H_T'\tilde{W}_TH_T\right)^{-1}H_T'\tilde{W}_Tg_{Tb}(\dot{b}_T)\right]\sqrt{T}(\hat{b}_T - b_0)\\ &= o_p(1). \end{aligned}\tag{A.15}$$

In the linear case,

$$\bar{b}_T = \hat{b}_T - \left(\bar{g}_2'\tilde{W}_T\bar{g}_2\right)^{-1}\bar{g}_2'\tilde{W}_T\left[\bar{g}_1 + \bar{g}_2\hat{b}_T\right] = \tilde{b}_T. \tag{A.16}$$

Proof of Proposition 1. For a positive definite matrix B and a conformable matrix A the nullspace of $A'(ABA')^-A$ equals the nullspace of A, by Lemma A.1, so that the conclusion follows with $A = LP_W$ and $B = V$.

Let P denote P_W for $W = V^{-1}$ and let $U = V^{-1}P$.

Proof of Proposition 2. Let F be a symmetric square root of V and let $N_1 = F^{-1}H(H'F^{-2}H)^{-1}H'F^{-1}$ and $N_2 = FP_W'(P_WF^2P_W')^-P_WF$. By Lemma A.4, N_1 and N_2 are idempotent, and by $P_WH = 0$, N_1 and N_2 are orthogonal. Furthermore, by Lemma A.6, $R(N_1) = q$ and $R(N_2) = r - q$, so that $R(N_1 + N_2) = R(N_1) + R(N_2) = r$. Since the only full rank, idempotent matrix is the identity, $I = N_1 + N_2$. If $R(Q) = r - q$, then $r - q = R(LP_W) = R(P_W)$, so that by Lemma A.4, $P_W'L'Q^-LP_W = P_W'(P_WVP_W')^-P_W = F^{-1}N_2F^{-1}$. Noting that $U = F^{-1}(I - N_1)F^{-1}$, the conclusion then follows by eq. (A.11) and the fact that $R(Q) = r - q$ implies

$$U - P'_W L' Q^- L P_W = F^{-1}(I - N_1 - N_2) F^{-1} = 0. \tag{A.17}$$

Numerical equality in the linear case follows from eq. (A.12).

Proof of Proposition 3. By $R([H, K]) = q + l$, $[H, K]' V^{-1} [H, K]$ is non-singular. Non-singularity of Q follows by partitioned inversion. Note that $LP = K' V^{-1} P = K'U$ and $Q = K'UK$, so that

$$\delta' K' P' L' Q^{-1} L P K \delta = \delta' K' U K Q^{-1} K' U K \delta = \delta' K' U K \delta. \tag{A.18}$$

The proof of Proposition 6 below implies that this non-centrality parameter is as large as possible in the class of GMM tests. By non-singularity of $Q = K'UK$ the non-centrality parameter is greater than zero for any non-zero δ, while by Proposition 1 any GMM test with less than l degrees of freedom will have a zero non-centrality parameter for some non-zero δ.

Proof of Proposition 4. Note that $H'\tilde{W}H$ non-singular follows from $[H, K]' V^{-1} [H, K]$ non-singular by partitioned inversion. Also note that from eq. (25) and partitioned inversion $(K' V^{-1} K)^{-1} \tilde{Q} (K' V^{-1} K)^{-1}$ is the lower right block of the inverse of $[H, K]' V^{-1} [H, K]$, which is also equal to $(K'UK)^{-1}$. Non-singularity of $\tilde{Q}$ follows immediately. Next, tedious but straightforward manipulation yields

$$L P_{\tilde{W}} = K' V^{-1} K (K'UK)^{-1} K'U. \tag{A.19}$$

Then eq. (A.19), $LP = K'U$, and $Q = K'UK$ imply that

$$P'_{\tilde{W}} L' \tilde{Q}^{-1} L P_{\tilde{W}} = U' K (K'UK)^{-1} K'U = P' L' Q^{-1} L P. \tag{A.20}$$

Asymptotic equivalence now follows from this equation and eq. (A.11). Numerical equivalence in the linear case follows from eq. (A.12).

Proof of Proposition 5. By Lemma 4, h_T is asymptotically equivalent (and numerically equivalent in the linear case) to a GMM test with $W = V^{-1}$ and $L = \tilde{L} = H'\tilde{W}$. By $H'U = 0$ and $UVU = U$,

$$\begin{aligned} \tilde{L} P V P' \tilde{L}' &= [\tilde{L} V - H'] U V U [V \tilde{L}' - H] \\ &= H' V^{-1} K (K' V^{-1} K)^{-1} K' U K (K' V^{-1} K)^{-1} K' V^{-1} H. \end{aligned} \tag{A.21}$$

By non-singularity of $K'UK$, $R(\tilde{L}PVP'\tilde{L}') = R(H'V^{-1}K)$. If $R(H'V^{-1}K) = l$, then by Lemma A.2 and eq. (A.21),

$$P'\tilde{L}'Q^{-}\tilde{L}P = U'K(K'UK)^{-1}K'U, \tag{A.22}$$

and asymptotic equivalence of h_T and $\tilde{m}_T$ follows by eqs. (A.11) and (A.20). Numerical equivalence in the linear case follows by eq. (A.12). If $R(H'V^{-1}K) < l$, then by Proposition 3.1 there is a non-zero δ such that the non-centrality parameter of h_T is zero, so that h_T will have smaller local power than the optimal tests for this δ. For $\delta = (K'V^{-1}K)^{-1}K'V^{-1}H\gamma$, for some γ such that δ is non-zero, the non-centrality parameter of h_T is equal to that of the optimal tests, so that for this δ, h_T has higher local power than the optimal tests because of its lower degrees of freedom.

Proof of Proposition 6. From eq. (A.17) it follows that the non-centrality parameter for an $r - q$ degree of freedom test is $\delta'K'UK\delta$. Also for any GMM test the proof of Proposition 2 implies $U = P_W'(P_W V P_W')^{-} P_W$. The conclusion follows from Lemma A.4, with $A = F^{-1}P_W'$ and $B = L'$.

References

Anderson, T.W. and H. Rubin, 1950, The asymptotic properties of estimates of the parameters in a single equation in a complete system of stochastic equations, Annals of Mathematical Statistics 21, 570–582.

Amemiya, T., 1973, Regression analysis when the dependent variable is truncated normal, Econometrica 41, 997–1016.

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.

Cumby, R.E., J. Huizinga and M. Obstfeld, 1983, Two-step two-stage least squares estimation in models with rational expectations, Journal of Econometrics 21, 333–355.

Domowitz, I. and H. White, 1982, Misspecified models with dependent observations, Journal of Econometrics 20, 33–58.

Eichenbaum, M.S., L.P. Hansen and K.J. Singleton, 1984, A time series analysis of representative agent models of consumption and leisure choice under uncertainty, Mimeo.

Engle, R.F., 1984, Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, Ch. 13 in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics (North-Holland, Amsterdam).

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, P.L., 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., 1984, Specification and estimation of simultaneous equations models, Ch. 7 in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics (North-Holland, Amsterdam).

Hausman, J.A. and W.E. Taylor, 1980, Comparing specification tests and classical tests, Manuscript (MIT, Cambridge, MA).

Hausman, J.A., W.K. Newey and W.E. Taylor, 1983, Efficient estimation and identification of simultaneous equations models with covariance restrictions, Department of Economics working paper 331 (MIT, Cambridge, MA).
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. (Université de Lausanne, Lausanne).
Lancaster, P., 1969, Theory of matrices (Academic Press, New York).
Mankiw, N.G., J. Rotemberg and L.H. Summers, 1982, Intertemporal substitution in macroeconomics, NBER working paper.
Newey, W.K., 1983, Specification testing and estimation using a generalized method of moments, Unpublished Ph.D. thesis (MIT, Cambridge, MA).
Pitman, E.J.G., 1949, Notes on non-parametric statistical inference (Columbia University, New York).
Rao, C.R., 1973, Linear statistical inference and its applications (Wiley, New York).
Rea, J.D., 1978, Indeterminancy of the Chow test when the number of observations is insufficient, Econometrica 46, 229.
Ruud, P., 1982, A score of consistency, Mimeo. (University of California, Berkeley, CA).
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., 1982, Instrumental variables regression with independent observations, Econometrica 50, 483–499.
White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations, Econometrica 52, 143–161.