SPECIFICATION TESTING AND ESTIMATION USING A GENERALIZED METHOD OF MOMENTS

by

WHITNEY KENT NEWEY

B.A. Brigham Young University
(1978)

Submitted to the
Department of Economics
in partial Fulfillment of the
Requirements of the
Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1983

Signature of Author ______________________________
Department of Economics
May 5, 1983

Certified by ______________________________
Thesis Supervisor

Accepted by ______________________________
Head Ph.D. Committee

SPECIFICATION TESTING AND ESTIMATION USING

A GENERALIZED METHOD OF MOMENTS

by

WHITNEY KENT NEWEY

Submitted to the Department of Economics
on May 5, 1983 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in Economics

ABSTRACT

Three theoretical applications of a generalized method of moments to econometrics are given. This generalized method of moments makes use of certain functions of the data and parameters which are specified a priori to have expectation zero at the true parameter value. The first application derives the statistical properties of model specification tests which are based on these functions of the data and parameters, and shows how these properties are useful for obtaining powerful diagnostic tests for parameter consistency in the linear simultaneous equations model. The second application considers specification testing for maximum likelihood estimation in a moments framework and suggests that tests based on optimal cross-products of exogenous variables and the maximum likelihood score are useful diagnostic tools for detecting parameter inconsistency. The third application uses this generalized method of moments to obtain strong necessary conditions for identification, efficient estimators, and powerful tests of overidentifying restrictions for the linear simultaneous equations sytsem with covariance matrix restrictions.

Thesis Supervisor: Jerry A. Hausman
Title: Professor of Economics

ACKNOWLEDGEMENTS

This thesis was written with the help and encouragement of many people. Jerry Hausman has been an inexhaustible source of ideas, suggestions and encouragement. In particular, Chapter 3 was the direct result of questions he posed. Dan McFadden suggested the general framework of Chapter 1, as well as the possibility of optimal specification tests as presented in Chapter 2. Jim Powell quickly and carefully read the entire thesis and his suggestions resulted in substantial improvement in its presentation and content. These three people constituted my thesis committee. In addition to the specific contributions mentioned above, each was available at almost any time to discuss the content and direction of this thesis.

I would like to thank participants of workshops at Brigham Young University, California Institute of Technology, Harvard University, Princeton University, Stanford University and Yale University for their helpful suggestions. I have also benefited from discussing my work with fellow graduate students, particularly Douglas Bernheim, David Runkle, Kenneth West and Michael Whinston.

Barbara did an excellent job of typing this thesis, which included suffering without complaint through major changes and typing hundreds of equations.

My deepest feelings of gratitude are reserved for my wife, Cheryl, and my children, Michael, Elisa and Jenniffer. Cheryl's care and support for me and our children was indispensible to completion of this thesis. My wife and children were also a source of joy which gave me constant motivation.

TABLE OF CONTENTS

Page No.

CHAPTER I

Generalized Method of Moments Specification Testing

I. Introduction

Most conclusions and predictions obtained by using econometric methods to summarize economic data depend crucially on model specification. The purpose of this paper is to extend and clarify specification testing methodology by considering the properties of a certain class of specification tests. We discuss local power properties of specification tests in a manner which allows comparison of different specification tests, including Hausman (1978) tests and conventional tests of overidentifying restrictions. We also clarify the relationship among specification tests based on fixed moment conditions, and derive new methods of obtaining moment conditions on which specification tests may be based.

Most econometric estimators are formed by making use of certain functions of the data and parameters which are specified to have expectation zero at the true parameter value. The specification that the expectation of these functions is zero at the true parameter value is referred to as orthogonality conditions. Most econometric estimators can be viewed as being obtained by setting sample moments of the orthogonality condition functions close to zero. For example, instrumental variables estimators are obtained by setting sample averages of cross products of residuals and instruments close to zero. When more orthogonality conditions than parameters are available, specification tests can be based on how close the sample moment orthogonality conditions are to zero when evaluated at the estimated parameter values. This is the type of specification test which we will refer to as a generalized method of moments (GMM) specification test.

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 inconsistent 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. The consequences of specifying that some of the orthogonality conditions remain uncontaminated under misspecification are examined. Since power properties of specification tests depend on the degrees of freedom, a method for obtaining the degrees of freedom of GMM specification tests is given. Section five gives two applications of the 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 $\underset{\sim}{z} = (z_1, z_2, \ldots)$ be a realization of a strictly stationary stochastic process, where z_t is an element of R^p. 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 x R^p to R^r. Define

$$(1.1) \quad g_T(\underset{\sim}{z},b) = \frac{1}{T} \sum_{t=1}^{T} g(z_t,b)$$

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

$$(1.2) \quad \min_{b \in B} g_T(\underset{\sim}{z},b)'W_T g_T(\underset{\sim}{z},b)$$

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 $g_T(\underset{\sim}{z},b)$ close to zero by minimizing the quadratic form $g_T(\underset{\sim}{z},b)'W_T g_T(\underset{\sim}{z},b)$. This class of estimators has been considered by Amemiya (1973) and Hansen (1982) among others. If b_o, $g(z,b)$ and the stochastic process for z satisfy the property

$$(1.3) \quad E(g(z,b_o)) = 0$$

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

(1.4) $E(g(z,b_o)) \neq 0$,

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

(1.5) $g_{Tb}(\underset{\sim}{z},\hat{b}_T)'W_T g_T(\underset{\sim}{z},\hat{b}_T) = 0$

where $g_{Tb}(\underset{\sim}{z},b) = \frac{\partial g_T}{\partial b}(\underset{\sim}{z},b)$, 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 $g_T(\underset{\sim}{z},\hat{b}_T)$ are to zero. For example, when doing instrumental variables estimation on an overidentified model, specification tests can be based on how close linear combinations of sample averages of cross products of instruments and residuals are to zero. These tests are what we refer to as GMM specification tests. As will become clear, this framework allows for unification of the asymptotic distribution theory under local misspecification of most specification tests.

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 $L_T g_T(\underset{\sim}{z},\hat{b}_T)$ are to zero, where the definition of closeness accounts 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 $\sqrt{T} L_T g_T(z, \hat{b}_T)$. Then the form of the GMM specification test statistic is

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

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 $E(g(z,b_0)) = 0$. 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 non-central 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 index the data by a misspecification parameter c which lies in R^u. For each n the distribution function of $(z_1,\dots,z_n)$ will be specified as $F_n(z_1,\dots,z_n,c)$, for each c. Where the expectation exists, define

$$(1.6)\quad h(b,c) = \int g(z,b)F_1(dz,c).$$

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,

$$(1.7)\quad h(b_o,c_o) = 0.$$

If $c = c_o$ we define the model to be correctly specified, since the orthogonality conditions hold in the population.

We can now allow for local misspecification as follows. Let $c_T = c_o + \delta/\sqrt{T}$. Assume that for each T, $(z_1,\dots,z_T)$ has a distribution function $F_T(z_1,\dots,z_T;c_T)$ for each sample size T. The situation is that the observations are drawn from a model for which the specification is incorrect for each T, but as T gets large the specification of the model drifts towards correctness at rate $\sqrt{T}$. This stochastic specification is not meant to represent any real world phenomenon. It is a much used device to allow for approximations to the power of asymptotic tests.

To see the implications of local misspecification for the asymptotic distribution theory, note that

(1.8) $E(\sqrt{T} g_T(\underset{\sim}{z}, b_o)) = \sqrt{T} h(b_o, c_T)$

and by expanding $h(b_o, c_T)$ around c_o

(1.9) $\lim_{T\to\infty} \sqrt{T} h(b_o, c_T) = \frac{\partial h}{\partial c}(b_o, c_o)\delta \equiv \alpha .$

Application of an appropriate central limit theorem gives the conclusion that $\sqrt{T} g_T(\underset{\sim}{z}, b_o)$ converges in distribution to a normally distributed random vector with mean α and variance

$$V = \lim_{T\to\infty} T\left[E(g_T(\underset{\sim}{z}, b_o) g_T(\underset{\sim}{z}, b_o)') - h(b_o, c_T) h(b_o, c_T)'\right]$$

In this set up we consider the hypothesis of correct specification to be $H_o: \alpha = 0$. When $\alpha = 0$, all of the orthogonality conditions are (locally) correctly specified, and when $\alpha \neq 0$, some orthogonality conditions are misspecified.

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

<u>Assumption 1.1</u>: 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 1.1 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 Chapter Two.

Assumption 1.2: The vector $g(z,b)$ is a measurable function on a measurable space Z, and for almost all $z \in Z$ a continuously differentiable function of b.

Assumption 1.3: 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 $\frac{\partial g_T(\underset{\sim}{z},b)}{\partial b}$ converges in probability to $\frac{\partial h}{\partial b}(b,c_o)$ uniformly in b on every compact subset of B.

Define

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

Assumption 1.4: The estimator $\hat{b}_T$ satisfies $\sqrt{T}[\partial g_T(\underset{\sim}{z},\hat{b}_T)/\partial b']W_T g_T(\underset{\sim}{z},\hat{b}_T) = o_p(1)$ for a sequence of matrices W_T satisfying plim $W_T = W$, W positive semi-definite and $H'WH$ non-singular. Also, $L_T - L = o_p(1)$ for L with rank $(L) = s$.

Assumption 1.5: The random vector $Y_T = \sqrt{T}(g_T(\underset{\sim}{z},b_o) - h(b_o,c_T))$

converges in distribution to a random variable $Y_o \sim N(0,V)$ where

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

and V is non-singular. Also $V_T \overset{p}{\to} V$.

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

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

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

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

<u>Assumption 1.6</u>: The sequence of generalized inverses Q_T^- satisfies $Q_T^- \overset{p}{\to} Q^-$ where Q^- is a generalized inverse of Q.

Assumption 6 is required because Q may be singular. A sufficient condition for Assumption 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.

It is useful to consider the special case which occurs when g(z,b) is linear in b, so that

(1.10) $g(z,b) = G_1(z) - G_2(z)b$

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 (1.2) is given by

(1.11) $b_T = (G_{2T}' W_T G_{2T})^{-1} G_{2T}' W_T G_{1T}$

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

Theorem 1.1. If $c_T = c_o + \delta/\sqrt{T}$ and assumptions 1.1-1.6 are satisfied then

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

converges in distribution to a non-central chi-squared distribution with degrees of freedom equal to rank (Q) and non-centrality parameter

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

Also, λ^2 does not depend on the choice of generalized inverse, and 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 an appendix.

An important property of GMM specification tests is that they are inconsistent against general forms of misspecification. This inconsistency for some specific specification tests has been noted in Bierens (1982). It is our purpose to show that this inconsistency is a fundamental phenomenon and is related to identification of parameters under misspecification.

The inconsistency which results can be illustrated in terms of the non-centrality parameter λ^2. Note that

(1.12) $P_W H = H - H(H'WH)^{-1} H'WH = 0$

Let $C(A)$ denote the column space of a matrix A. Then for any α in $C(H)$, $\lambda^2 = 0$. The non-centrality parameter is zero on at least a q dimensional linear subspace of R^r. In fact, it is the case that the subspace of R^r on which λ^2 is zero an r-d dimensional subspace of R^r where d = rank Q is the degrees of freedom of the GMM specification test under consideration.

<u>Proposition 1.2</u>. If rank (Q) = d, then the set of α such that

$\lambda^2 = \alpha' P_W' L' Q^- L P_W \alpha = 0$ is an r-d dimensional subspace of R^r.

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 non-local 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 assumptions, in addition to Assumptions 1.1-1.6.

Assumption 1.7. 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, $g_T(\underset{\sim}{z},b)$ 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 1.8. 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 1.9. $L(c)$ and $V(c)$ are once continuously differentiable functions of c in C.

Assumption 1.10. For each c in C, $\sqrt{T}(L_T - L(c)) = 0_p(1)$ and $\sqrt{T}(W_T - W(c)) = 0_p(1)$.

Assumption 1.11. For each c in C, $\sqrt{T}(g_T(\underset{\sim}{z},b(c)) - h(b(c),c)) = 0_p(1)$

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

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 1.12. 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 non-singular.

Theorem 1.3. If Assumptions 1.1-1.2 are satisfied, then there is an open set N in R^u containing c_o such that the set of c in N satisfying $m_T = O_p(1)$ is a u-s dimensional C^1 sub-manifold of N.

Note that Assumption 1.12 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 verifies the claim 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 inconsistency 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

$$G(z,b,c) = g(z,b) - h(b,c).$$

If h(b,c) were known, estimation of b and c could be attempted using the orthogonality condition function G(z,b,c). But now there are q + u parameters to be estimated using r orthogonality conditions, and if q + u > r or u > r - q, then an order condition fails and then b and c are not identified from the orthogonality condition function G(z,b,c). The result that b and c are not identified under the alternative results in GMM specification tests being inconsistent, since the value of $\bar{c}$ cannot be consistently estimated.

Several other comments should be made concerning Theorem 1.3. The inconsistency of GMM specification tests is quite distinct from the inconsistency of the Hausman test discussed in Holly (1982). In Holly's framework, a consistent specification test was available due to a maintained hypothesis. We discuss the implications of maintained hypotheses in Section Four. Also, it is important to note that Theorem 1.3 only implies that specification tests using fixed orthogonality conditions are inconsistent when no maintained hypotheses are imposed. Non-nested may be are consistent due to a maintained hypothesis concerning the form of an alternative model. Also, a specification test for non-linear regression which is consistent against general forms of misspecification is presented in Bierens (1982). Bierens allows for orthogonality condition functions which depend on the alternative through the data.

Finally, it is an interesting fact that the set of c values for which the GMM specification test fails is of measure zero in R^p. If a prior distribution was specified for c which has a density at every $c \neq c_o$, the prior probability that a GMM test fails is zero. Of course the test fails with probability one if the c parameter generating the data lies in the inconsistency manifold. However, we conjecture that for a given c parameter, the set of orthogonality condition functions giving a consistent specification test is large in an appropriate sense. These two considerations suggest that the failure of GMM tests for large samples will occur very rarely in applications.

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

$$(1.13)\quad \min_b \; g_T(\underset{\sim}{z},b)'A_T g_T(\underset{\sim}{z},b)$$

and $\bar{b}_T$ from

$$(1.14)\quad \min_b \; g_T(\underset{\sim}{z},b)'C_T g_T(\underset{\sim}{z},b)$$

To assumptions 1.1-1.7 we add the following assumptions

Assumption 1.13. $\bar{b}_T \xrightarrow{p} b_o$ and $\tilde{b}_T \xrightarrow{p} b_o$.

Assumption 1.14. $A_T \xrightarrow{p} A$ 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 $\sqrt{T}\, \partial g_T(\underset{\sim}{z},\bar{b}_T)/\partial b'\; C_T\; g_T(\underset{\sim}{z},\bar{b}_T) = o_p(1)$

and $\sqrt{T}\,\partial g_T(\underset{\sim}{z},\tilde{b}_T)/\partial b' A_T g_T(\underset{\sim}{z},\tilde{b}_T) = o_p(1)$.

Let

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

and

$$\begin{aligned} M = {} & (H'CH)^{-1} H'CVCH(H'CH)^{-1} + (H'AH)^{-1} H'AVAH(H'AH)^{-1} \\ & - (H'CH)H'CVAH(H'AH)^{-1} - (H'AH)^{-1} H'AVCH(H'CH)^{-1} \end{aligned}$$

The matrix M will be the asymptotic covariance matrix of q_T. To obtain a consistent estimator of M_T define the matrices

$$\tilde{H}_T = g_{Tb}(\underset{\sim}{z},\tilde{b}_T) \quad , \quad \bar{H}_T = g_{Tb}(\underset{\sim}{z},\bar{b}_T)$$

and let

$$\begin{aligned} M_T = {} & (\bar{H}_T' C_T \bar{H}_T)^{-1} \bar{H}_T' C_T V_T C_T \bar{H}_T (\bar{H}_T' V_T \bar{H}_T)^{-1} \\ & + (\tilde{H}_T' A_T \tilde{H}_T)^{-1} \tilde{H}_T' A_T V_T A_T \tilde{H}_T (\tilde{H}_T' A_T \tilde{H}_T)^{-1} \\ & - (\bar{H}_T' C_T \bar{H}_T)^{-1} \bar{H}_T' C_T V_T A_T \tilde{H}_T (\tilde{H}_T' A_T \tilde{H}_T)^{-1} \\ & - (\tilde{H}_T' A_T \tilde{H}_T)^{-1} \tilde{H}_T' A_T V_T C_T \bar{H}_T (\bar{H}_T' A_T \bar{H}_T)^{-1} \end{aligned}$$

where V_T is positive definite.

<u>Assumption 1.15</u>. The sequence of g-inverses M_T^- is chosen so that plim $M_T^- = M^-$.

As with Assumption 1.6, Assumption 1.15 is required due to possible singularity of M. Note that in the form we have given the estimator of M given by M_T will be constrained to be positive semi-definite.

Now define the Hausman test statistic

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

<u>Theorem 1.4</u>. If Assumptions 1.1-1-6 and 1.13-1.15 are satisfied then h_T converges in distribution to a non-central 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 λ_h^2 does not depend on the choice of generalized inverse and if g(z,b) is linear in b h_T does not depend on the choice of g-inverse M_T^-.

It should be emphasized that Theorem 1.4 gives the asymptotic distribution under local misspecification of most of the Hausman tests which have been proposed in the literature. Further, Theorem 1.4 allows for comparison of first-order power properties of these different Hausman for specific forms of misspecification. Section four gives some specific comparisons. The generality of Theorem 1.4 is a consequence of the fact that most econometric estimators can be viewed as GMM estimators. Most instrumental variable and maximum likelihood estimators fit into this framework. 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 1.4. Note also that the form of misspecification we consider includes various forms of correlation of variables with residuals and also includes finitely parameterized likelihood function misspecification.

An attraction of the Hausman test, as presented in Hausman (1978) is its computational simplicity. In particular, the use of an asymptotically efficient estimator in forming q_T implies that the asymptotic variance of q_T is just the difference of the variances of the inefficient estimator and the efficient estimator. This type of simplification is also present in the GMM estimation framework. Note that if $C = V^{-1}$, then

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

which is the difference of the asymptotic variance of $\tilde{b}_T$ and $\bar{b}_T$. As shown in Hansen (1982), when $C = V^{-1}$, $\bar{b}_T$ is asymptotically efficient in the class of GMM estimators arising from use of orthogonality condition functions $g(z,b)$, or in other words, V^{-1} is the efficient choice of weighting matrix W. It can be shown that this result is an explanation of why all of the specification tests discussed in Hausman (1978) have the simple matrix difference form even if the disturbances are not normally distributed. An estimator of M when $C=V^{-1}$, which is constrained to be positive semi-definite, is given by

(1.16) $M_T = (\bar{H}_T'A_T\bar{H}_T)^{-1}\bar{H}_T'A_TV_TA_T\bar{H}_T(\bar{H}_T'A_T\bar{H}_T)^{-1} - (\bar{H}_T'V_T\bar{H}_T)^{-1}$.

There is a simple first-order 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

(1.17) $\dot{b}_T = \tilde{b}_T - (\tilde{H}_T'C_T\tilde{H}_T)^{-1}\tilde{H}_T'C_Tg_T(\underset{\sim}{z},\tilde{b}_T)$.

Theorem 1.5. If Assumptions 1.1-1.6 and 1.13-1.15 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 1.5 is the appropriate generalization to the GMM framework of the well-known one-step theorems for maximum likelihood and non-linear least squares, and holds when the model is locally misspecified.

Theorem 1.5 implies that

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

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

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

(1.19) $W = A \ , \ L = H'C.$

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 asymptotically equivalent to a GMM specification test with

(1.20) $W = C \ , \ L = H'A$

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

Theorem 1.5 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 non-centrality parameter λ_h^2 of Theorem 1.4 will be zero on a subset of α values in R^r which is an $r-d_h$ dimensional linear subspace of R^r. A result exactly analogous to Theorem 1.3 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 non-central chi-squared distributions. The tail probability of a non-central chi-squared distribution is increasing in the non-centrality parameter and decreasing in the numbers of degrees of freedom. Therefore, the local power of a GMM specification test increases with the non-centrality parameter and decreases with the degrees of freedom for a fixed asymptotic significance level. The degrees of freedom of a GMM specification test play a central role in determining first order asymptotic power and knowledge of the degrees of freedom is essential for determining the correct asymptotic significance level of these tests. The following two results give a convenient method of determining degrees of freedom of specification tests when the asymptotic covariance matrix, V, of $\sqrt{T}g_T(\underset{\sim}{z},b_o)$ is non-singular. Let R(A) denote the rank of the matrix A.

Proposition 1.6. If V and H'WH are non-singular then

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

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

Corollary 1.7. If V, H'AH and H'CH are non-singular then

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

Recall that Q is the asymptotic covariance of the sample moments $L_T g_T(z, \hat{b}_T)$ and M is the asymptotic covariance matrix of the difference of two estimators $q_T = \tilde{b}_T - \bar{b}_T$. 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.

It is of some interest to know if there is a general class of asymptotically equivalent GMM specification tests. Somewhat surprisingly, there is such a class of 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 1.8. If Assumptions 1.1-1.6 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 $LP_W V P_W' L'$. 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 test 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 1.8 is that any GMM test with the maximum number of degrees of freedom, r-q, is asymptotically equivalent to any test with degrees of freedom r-q and numerical equality holds if g(z,b) is linear in b. Note that we have implicitly assumed that the same etimator V_T of V is used in forming each test statistic. Our results on numerical equality of various test statistics depend on this assumption.

In order to distinguish different specification tests on the basis of first-order asymptotic power, the degrees of freedom considered must be less than r-q. As a benchmark for comparison, it is useful to know the maximum value the non-centrality parameter can have.

Lemma 1.9. The non-centrality 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 non-centrality 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

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

by Proposition 1.6, and by Theorem 1.1 has non-centrality 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

$$(1.21)\quad E(g(z,b_o)) = \bar{c}$$

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 kxl vector, with $k \leq r - q$. Let $U = V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}V^{-1}$. Choosing $W = V^{-1}$ and defining $P = P_W$, the maximum value for the non-centrality 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

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

The fact that the order condition for identification under misspecification is satisfied suggests that U_{22} should be non-singular. Partition H, V and V_T conformably with $\alpha = (0,\alpha_2')'$,

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

<u>Lemma 1.10.</u> If V is non-singular and H_1 has rank q, then U_{22} is non-singular.

The condition that H_1 has rank q is just a rank condition for identification of b_o and $\bar{c}$ under misspecification: see chapter 3. The non-singularity of U_{22} implies that λ^{*2} equals zero if and only if $\alpha_2 = 0$.

The maximum value of λ^{*2} is attained by any GMM specification test with degrees of freedom r-q. If $k = r - q$, so that the parameters are exactly identified under misspecification, the only consistent tests will be GMM tests with degrees of freedom r-q, which are all asymptotically equivalent by Proposition 1.8. When $k < r - q$ the model is overidentified under misspecification. There are the $q + k$ elements of $(b,\bar{c}_2)$, where $\bar{c}_2$ are the non-zero elements of $\bar{c}$, which are parameters which have to be estimated, and there are $r > q + k$ orthogonality conditions with which to estimate the parameters. Since the model is overidentified under misspecification, a consistent test with higher local power than the r-q degrees of freedom test can be obtained by of freedom k and non-centrality parameter λ^{*2}. One such optimal test can be obtained as follows.

<u>Proposition 1.11</u>. If Assumptions 1.1-1.6 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 non-central chi-squared distribution with k degrees of freedom and non-centrality parameter $\lambda^{*2} = \alpha_2' U_{22}\alpha_2$.

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

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

partitioned conformably with α,

$$(1.22)\quad \overline{m}_T = [g_{T2}(\underset{\sim}{z},\hat{b}_T)' - g_{T1}(\underset{\sim}{z},\hat{b}_T)'V_{11T}^{-1}V_{12T}]Q_T^{-1}$$
$$[g_{T2}(\underset{\sim}{z},\hat{b}_T) - V_{21T}V_{11T}^{-1}g_{T1}(\underset{\sim}{z},\hat{b}_T)],$$

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

$$\min_b g_T(\underset{\sim}{z},b)'V_T^{-1}g_T(\underset{\sim}{z},b)$$

Since b_o is overidentified, it is the case that if $V_{12} \neq 0$ (i.e., $g_{T1}(\underset{\sim}{z},b_o)$ is correlated with $g_{T2}(\underset{\sim}{z},b_o)$) then the optimal test of whether $E[g_{T2}(\underset{\sim}{z},b_o)] = 0$ combines the sample moments $g_{T1}(\underset{\sim}{z},\hat{b}_T)$ with $g_{T2}(\underset{\sim}{z},\hat{b}_T)$, via the linear combination matrix L_T. This test statistic has also been independently derived in Eichenbaum, Hansen, and Singleton (1982), in a somewhat different context. They view the non-zero elements of $\overline{c}$ as parameters to be estimated, and derive Wald, gradient and Gallant and Jorgenson (1979) forms of test statistics for the hypothesis that the non-zero elements of $\overline{c}$ are zero. The statistic $\overline{m}_T$ of equation (1.22) is a gradient statistic, since all the orthogonality condition functions are used to obtain $\hat{b}_T$. Their Wald test is also a GMM test with L given as in Proposition 1.11.

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 $\hat{b}_T$ and the GMM estimator $\tilde{b}_T$ which is obtained by solving

$$(1.23) \quad \min_b \; g_{1T}(\underset{\sim}{z},b)'V_{11T}^{-1} \; g_{1T}(\underset{\sim}{z},b).$$

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 Hausman specification test statistic be given by

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

<u>Proposition 1.12</u>. If Assumptions 1.1-1.6 and 1.12-1.15 are satisfied for the Hausman test based on $q_T = \tilde{b}_T - \hat{b}_T$ then $\bar{h}_T$ converges in distribution to a non-central chi-squared distribution with degrees of freedom

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

and non-centrality parameter

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

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

Proposition 1.13. 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 1.3 implies that this test is inconsistent against general forms of misspecification, with direction of inconsistency for plim $g_T(b_o) \neq 0$ given by the linear space spanned by the columns of $E[\partial g_T(b_o)/\partial b]$. Also, Proposition 1.8 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. For $g(z,b)$ linear in b, the test statistics are invariant with respect W_T. Note, though, that we have implicitly assumed that the same estimator V_T of the asymptotic covariance matrix V is used to form each test statistic. In practice, the test statistics could differ in the linear case due to different choices of V_T. In each application, our statements concerning numerical equality of statistics will be subject to the caveat that the same V_T be used for each.

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 $\text{rank}[AH\,\vdots\, WH] = r$. Then for $\tilde{b}_T$ obtained from equation (1.13) Corollary

1.7 implies the Hausman test based on the difference $\hat{b}_T - \tilde{b}_T$ has degrees of freedom r-q. By Proposition 1.8 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 chapter 3.

Our second application concerns tests of overidentifying restrictions in the linear simultaneous equations system. Note, first, that our preceding discussion has immediate implications for linear system specification tests. In particular, it follows from Proposition 1.8 that the system test of overidentifying restrictions based on the Gallant and Jorgenson (1979) criteria for three-stage least-squares (3SLS) (which is just a minimum chi-square test, as discussed in Hansen (1982)) is numerically equal to the specification test based on the difference of the two-stage least-squares (2SLS) and 3SLS 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

$$(1.29)\quad y_1 = Y_1\beta + Z_1\delta + u = X_1 b + u,$$

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 TxK 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 heteroskedasticity, so that at the true parameter vector

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

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

(1.26) $\min_b \; u'Z(Z'Z)^{-1}Z'u.$

Note that by equation (1.25) $\hat{b}_T$ is an optimal GMM estimator obtained from the orthogonality condition functions $Z_t'u_t = Z_t'(y_{1t} - X_{1t}b)$. Let $\hat{u} = y_1 - X_1 \hat{b}_T$ and $\hat{\sigma}^2 = \hat{u}'\hat{u}/T$. 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

$$\tilde{m}_T = \frac{\hat{u}'Z}{\sqrt{T}} V_T^{-1} \frac{Z'\hat{u}}{\sqrt{T}} = T \frac{\hat{u}'Z(Z'Z)^{-1}Z'\hat{u}}{\hat{u}'\hat{u}} ;$$

see Sargan (1980) or Hansen (1982). In the terminology of our paper the statistic $\tilde{m}_T$ is the GMM specification test statistic with maximum degrees of freedom K-q. As noted by Hausman (1982),

(1.27) $\tilde{m}_T = T\tilde{R}^2$

where $\tilde{R}^2$ is the r-squared from a regression of $\hat{u}$ on Z.

For the statistic $\tilde{m}_T$, the set of alternatives against which the

GMM test statistic is inconsistent can be easily interpreted. Since the orthogonality condition functions are linear in b, $\tilde{m}_T = O_p(1)$ if and only if plim $Z'u_o/T = H(-\alpha)$, for $u_o = y_1 - X_1 b_o$, b_o the true parameter vector, $H = \text{plim}(-Z'X_1/T)$ and α is a Kxq vector. The test of overidentifying restrictions fails when u_o is the sum of an uncontaminated component and a linear combination of right-hand side variables. Define $\eta = u_o - X_1\alpha = u_o - Y_1\alpha_1 - Z_1\alpha_2$. Then plim $Z'\eta/T = 0$ and $y_1 = Y_1(\beta_o + \alpha_1) + Z_1(\gamma_o + \alpha_2) + \eta$ so that plim $b = b_o + \alpha$. Intuitively, the 2SLS coefficients absorb all misspecification and none is left over to be detected by $\tilde{m}_T$.

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 than that based on $\tilde{m}_T$ are available. Since these tests will be based on the 2SLS estimator $\hat{b}_T$ they will be gradient (i.e., Lagrange multiplier) tests.

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

(1.28) $y_1 = Y_1\beta + Z_1\gamma + WC + \varepsilon$

where $W = [\tilde{Y}_2, \tilde{Z}_2]$ consists of $\ell \leq K-q$ columns of $[Y_2 Z_2]$. Following Burgette, Gallant and Souza (1981) we can define a gradient test of H_o: $C=0$ using the criterion function of the minimization problem (1.26). For ε defined implicitly in equation (1.28) we have

(1.29) $\partial/\partial C[\varepsilon' Z(Z'Z)^{-1} Z'\varepsilon] = - W'Z(Z'Z)^{-1} Z'\varepsilon$

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

(1.30) $[V_T^{-1} H_T \vdots L_T'] = [\sigma^2 (Z'Z)^{-1} Z'X_1 \vdots (Z'Z)^{-1} ZW]$

so that by Proposition 1.6 the degrees of freedom of this test will be ℓ so long as plim $Z'[X_1 \vdots W]/T$ has rank $g+\ell$. Define $\hat{X}_1 = Z(Z'Z)^{-1} Z'X_1$ and $\hat{W} = Z(Z'Z)^{-1} Z'W$. In terms of our previous notation, $H_T = - Z'X_1/T$ so that

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

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

$$(1.32)\quad m_T = u'ZL_T'[L_T Z'Q_{X_1} ZL_T']^{-1} L_T Z'u\cdot(1/\sigma^2)$$
$$= Tu'W(W'Q_{X_1} W)^{-1} W'u/u'u,$$

where the non-singularity of the matrix $W'Q_{X_1}W$ for large T follows from the degrees of freedom of m_T being equal to ℓ. Note that it follows from $X_1'u = 0$ that $m_T = TR^2$ where R^2 is the r-squared from a regression of u on $[X_1, W]$.

While we have not explicitly considered the power properties of tests of parametric hypotheses such as H_0: C = 0, the following direct argument shows that m_T gives an optimal GMM test. Let C_T satisfy $\lim \sqrt{T}C_T = C_0$. Then by $u_0 = WC_T + \varepsilon$, $Z'u_0/\sqrt{T} \xrightarrow{d} N(\text{plim}(Z'W/T)C_0, V)$ so that

$$(1.33)\quad \lambda^{*2} = C_0' \,\text{plim}(1/T\sigma^2)[W'Z(Z'Z)^{-1}$$
$$- (Z'Z)^{-1} Z'X_1(X_1'X_1)^{-1}X_1'Z(Z'Z)^{-1}]Z'WC_0$$
$$= C_0' \,\text{plim}(1/T\sigma^2)(W'Q_{X_1}W)C_0$$

Further, since $L_T P_T = W'Z(Z'Z)^{-1}[I_K - Z'X_1(X_1'X_1)^{-1}X_1'Z(Z'Z)^{-1}]$, the non-centrality parameter of m_T is given by

$$(1.34)\quad \lambda^2 = C_0' \,\text{plim}[(1/T\sigma^2)W'ZP_T'L_T'(W'Q_{X_1}W)^{-1}L_T P_T Z'W]$$
$$= C_0'\text{plim}[(1/T\sigma^2)W'Q_{X_1}W(W'Q_{X_1}W)^{-1}W'Q_{X_1}W]C_0 = \lambda^{*2}$$

The second particular form of misspecification which is of interest has some instrumental variables correlated with u. For this form of misspecification we can use the optimal GMM test of Proposition 1.11. Suppose that $Z = [Z^1, Z^2]$ where Z^2 is a Txm vector, $m < K-q$, which is correlated with u_o when misspecification is present and Z^1 is a Tx(K-m) vector of predetermined variables which remain uncorrelated with u_o. Let $H_1 = -\operatorname{plim} Z^{1\prime}X_1/T$ have rank q, which guarantees this test will have m degrees of freedom. Then from Proposition 1.11 this optimal GMM test has $L_T = [-Z^{2\prime}Z^1(Z^{1\prime}Z^1)^{-1}, I_m]$. Note that $ZL_T' = \tilde{V} = (I_T - Z^1(Z^{1\prime}Z^1)^{-1}Z^{1\prime})Z^2$ is the Txm matrix of residuals from a regression of Z^2 on Z^1. The test statistic is given by

$$(1.35)\quad \bar{m}_T = \hat{u}'ZL_T'(L_TZ'Q_{\hat{X}_1}ZL_T')^{-1}L_TZ'\hat{u}/\hat{\sigma}^2$$

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

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 derived in Hausman and Taylor (1980). By Proposition 1.12 this Hausman test has degrees of freedom given by

(1.36) $\bar{d}_h = \text{rank}(\text{plim } Z'[Z^1, X_1]/T) + m-K$

As discussed in Hausman and Taylor (1980), the rank of the matrix plim $Z'[Z^1, X_1]/T$ is equal to $\min(K, q+K-m-r)$, where r is the number of common columns of Z^1 and $X_1 = [Y_1, Z_1]$. Then $\bar{d}_h = \min(m, q-r)$, where r is also the number of common elements of Z^1 and Z_1. By Theorem 1.5 and linearity in b the Hausman test statistic is equal to a GMM test with $L_T = [X_1'Z^1(Z^{1\prime}Z^1), 0]$. 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 $S'\hat{\hat{X}}_1'Q_{\hat{X}_1}\hat{\hat{X}}_1S'$ is non-singular. Then by Lemma A2 of the appendix

(1.37) $\bar{h}_T = \hat{u}'ZL_T'(L_TZ'Q_{\hat{X}_1}ZL_T)^- L_TZ'\hat{u}/\hat{\sigma}^2$

$= u'\hat{\hat{X}}_1(\hat{\hat{X}}_1'Q_{\hat{X}_1}\hat{\hat{X}}_1)^-\hat{\hat{X}}_1'\hat{u}/\hat{\sigma}^2$

$= T\hat{u}'\hat{\hat{X}}_1S(S'\hat{\hat{X}}_1'Q_{\hat{X}_1}\hat{\hat{X}}_1S)^{-1}S'\hat{\hat{X}}_1'\hat{u}/\hat{u}'\hat{u}$

$= T\bar{R}_h^2$

where $\bar{R}_h^2$ is the R-squared from a regression of u on $[\hat{X}_1, \hat{\hat{X}}_1S]$. Proposition 1.13 implies that if $d_h = m$, then $\bar{h}_T = \bar{m}_T$.

When a full set of overidentifying restrictions is being tested then each of these test statistics is identical. That is, if $\ell = m = \bar{d}_h = K-q$, then Proposition 1.8 implies that $\tilde{m}_T = m_T = \bar{m}_T = \bar{h}_T$. Note that these equalities hold <u>independently</u> of the particular set of overidentifying coefficient restrictions being tested and 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 should be used. However, even when a particular form of misspecification occurs, the statistic $\tilde{m}_T$ retains a certain kind of robustness. Its non-centrality parameter is at least as large as that of the individual statistics, no matter what form the misspecification takes. Power loss from use of $\tilde{m}_T$ rather than a specific statistic will result from $\tilde{m}_T$ having larger degrees of freedom. If the degrees of freedom difference is small, then this power loss may also be small.

VI. Conclusions.

We have presented results for a class of specification tests which we have referred to as generalized method of moments specification tests, and shown that Hausman tests are first-order equivalent to these tests. Due to lack of identification under general misspecification, GMM specification tests will be inconsistent. 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 general test of overidentifying restrictions given in Hansen (1982).

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 nx1 vector function of x and b such that, if the model is correctly specified, the conditional expectation $E[e(z_t,b_o)|I_t]$ satisfies

(1.38) $E[e(z_t,b_o)|I_t] = 0,$

where I_t is a conditionin 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

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

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 the next chapter 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 to Chapter One

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

$$\text{(A1)}\quad (A(A'A)^-A' - AB(B'A'AB)^-B'A')^2$$
$$= A(A'A)^-A' - A(A'A)^-A'AB(B'A'AB)^-B'A'$$
$$- AB(B'A'AB)^-B'A'A(A'A)^-A'$$
$$+ AB(B'A'AB)^-B'A'AB(B'A'AB)^-B'A'$$
$$= A(A'A)^-A' - AB(B'A'AB)^-B'A'.$$

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

<u>Lemma A5</u>: 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, $R(AB) =$

$R([C \vdots B]) - n$.

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 $[C \vdots B]\, [-z',x']' = 0$. Similarly, suppose $[C \vdots B]\, [-z',x']' = 0$. Then $Cz = Bx$ implies $ABx = ACz = 0$. Therefore $N(AB)$ is isomorphic to $N([C \vdots B])$ and so $\dim N(AB) = \dim N([C \vdots B])$. By Lancaster (1969) Theorem 1.16.2 $R(AB) = m - \dim N(AB)$ and $R([C \vdots B]) = m + n - \dim N([C \vdots B])$, so that $R(AB) - m = R[C \vdots B] - m - n$ or $R(AB) = R[C \vdots B] - n$.

<u>Lemma A6</u>: 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 1.1: 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

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

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

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

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

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

By Assumption 1.3, 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

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

and

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

by the usual rules for probability limits of sums and products of random variables. By Assumption 1.5 and the definition of h(b,c),

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

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

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

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

where $|\tilde{c}_T - c_o| \leq |c_T - c_o|$. By Assumption 1.3 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)

(A.8) $\sqrt{T}\, g_T(b_o) = Y_T + \alpha + o(1) = O_p(1).$

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

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

Then since H'WH is non-singular by Assumption 1.4, $(g_{Tb}(\hat{b}_T)'W_T g_{Tb}(\tilde{b}_T))^{-1}$ exists with probability approaching one, and by equations (A.9) and (A.4)

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

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

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

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

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 $\sqrt{T}\ g_T(b_T) = O_p(1)$,

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

by Assumption 1.5. By Assumption 1.6, equation (A.12) and $\sqrt{T}\ L_T g_T(\hat{b}_T) = O_p(1)$,

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

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

from which follows the fact that m_T converges in distribution to a non-central chi-squared with degrees of freedom rank (Q) and non-centrality parameter λ^2. The invariance of λ^2 with respect to choice of g-inverse follows from Lemma A3 and Lemma A6.

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

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

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

so that

(A.15) $m_T = TG'_{1T}P'_{WT}L'_T(L_TP_{WT}V_TP'_{WT}L'_T)^{-}L_TP_{WT}G_{1T}$,

and invariance of m_T follows by the same argument as invariance of λ^2 with respect to g-inverse Q^-.

Proof of Proposition 1.2: Recall that $Q = LP_wVP'_wL'$. Since $\dim N(P'_wL'Q^-LP_w) = r - R(P'_wL'Q^-LP_w)$ by Lancaster (1969) Theorem 1.16.2, it suffices to show that

(A.16) $R(P'_wL'Q^-LP_w) = R(Q)$

Equation (A.14) follows immediately upon application of Lemma A6.

Proof of Theorem 1.3: 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

(A.17) $\frac{\partial h}{\partial b}(b(c),c)'W(c)h(b(c),c) = 0$.

By the implicit function theorem, $h(b_o, c_o) = 0$ and $H_o'WH_o$ non-singular, equation (A.17) implies

$$(A.18)\ \frac{\partial b}{\partial c}(c_o) = -(H_o'W_oH_o)^{-1}H_o'W_oK_o.$$

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 1.12, $\frac{\partial J}{\partial c_o}$ 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.19). By J(c) continuous on N', J(c) has rank s on for all c in an open neighborhood N'cN'. 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 $m_T = 0_p(1)$.

Let $f_T(\underset{\sim}{z}, b) = 1/2\ g_T(\underset{\sim}{z}, b)'W_Tg_T(\underset{\sim}{z}, b)$. By Assumption 1.7, $f_T(\underset{\sim}{z}, b)$ converges in probability to f(b,c) uniformly in b. Assumption 1.8 and a convergence in probability version of Amemiya (1973) Lemma 3 (such as that given in McFadden (1980)) implies that there exists a measureable b_T solving

$$(A.20)\quad \min_{b \in B}\ f_T(\underset{\sim}{z}, b)$$

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

(A.21) $g_{Tb}(\underset{\sim}{z},\hat{b}_T)'\, W_T \sqrt{T}\, g_T(\underset{\sim}{z},\hat{b}_T) = o_p(1)$.

For notational convenience, the $\underset{\sim}{z}$ argument will be suppressed from now on. Expanding $g_T(\hat{b}_T)$ in a mean value expansion (as in the proof of Theorem 1.1) equation (A.18) implies

(A.22) $g_{Tb}(\hat{b}_T)'W_T g_{Tb}(\bar{b}_T)\sqrt{T}\,(\bar{b}_T - b(c))\, +$

$$g_{Tb}(\hat{b}_T)W_T \sqrt{T}\, g_T(b(c)) = o_p(1)$$

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.17)

(A.23) $g_{Tb}(\hat{b}_T)'W_T\sqrt{T}\, g_T(b(c)) = g_{Tb}(\hat{b}_T)'W_T\sqrt{T}\,(g_T(b(c)) - h(b(c),c))$

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

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

By Assumption 1.17, and plim $\hat{b}_T = b(c)$, plim $g_{Tb}(\bar{b}_T) = H(c)$. Then by Assumptions 1.10 and 1.11 the first two terms after the equality in equation (A.23) are $O_p(1)$. By Assumption 1.7 we can apply a mean-value expansion to the last term to obtain

(A.24) $\sqrt{T}[g_{Tb}(\hat{b}_T) - H(c)] = \sqrt{T}[g_{Tb}(b(c)) - H(c)]$

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

with $|\bar{b}_T - b(c)| \leq |b_T - b(c)|$. Then by Assumption 1.1 and equations (A.23) and (A.24)

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

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

By Assumption 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, $H_o' W_o H_o$ non-singular and $h(b_o,c_o) = 0$, it follows that there is N < N" such that

(A.25) $\sqrt{T}(\hat{b}_T - b(c)) = O_p(1)$.

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

$J(c) = L(c)h(b(c),c) = 0$,

$$(A.26)\quad L_T\sqrt{T}\, g_T(\hat{b}_T) = L_T g_{Tb}(\tilde{b}_T)\sqrt{T}\,(b_T - b_o)$$
$$+ \sqrt{T}[L_T - L_T(c)]g_T(b(c))$$
$$+ L(c)\sqrt{T}[g_T(b(c)) - h(b(c),c)].$$

The first and second terms are $O_p(1)$ by equation (A.25), its proof, Assumption 1.10, and Assumption 1.7 which implies plim $g_T(b(c)) = h(b(c),c)$ so that $g_T(b(c)) = O_p(1)$. By Assumption 1.11, the second term is also $O_p(1)$.

Proof of Theorem 1.4: 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

$$(A.27)\quad \sqrt{T}(\tilde{b}_T - b_o) = -\,(H'A'H)^{-1}HA(Y_T + \alpha) + o_p(1)$$

and

$$(A.28)\quad \sqrt{T}(\bar{b}_T - b_o) = -\,(H'CH)^{-1}H'C(Y_T + \alpha) + o_p(1).$$

Subtraction yields

$$(A.29)\quad \sqrt{T}\, q_T = [(H'CH)^{-1}H'C-(H'AH)^{-1}H'A](Y_T + \alpha) + o_p(1).$$

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

$$(A.30)\quad \sqrt{T}\, q_T \overset{d}{\rightarrow} D(Y_o + \alpha) \sim N(D\alpha,\ DVD').$$

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

$$(A.31)\quad h_T = Tq_T'(M_T^- - M^-)q_T + Tq_T'M^-q_T$$

$$= Tq_T'M^-q_T + o_p(1)$$

and the asymptotic distribution result follows. To see that λ_h^2 is independent of the choice of g-inverse, note that

$$\lambda_h^2 = \alpha'D'(DVD')D^-D\alpha$$

and the conclusion follows by the same argument as used in the proof of Theorem 1.1.

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

$$(A.32)\quad \tilde{b}_T - \bar{b}_T = - D_T G_{1T}$$

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

Further,

$$(A.33)\quad h_T = T(\tilde{b}_T - \bar{b}_T)'(D_TV_TD_T')^-(\tilde{b}_T - \bar{b}_T)$$

$$= TG_{1T}'D_T'(D_TV_TD_T')^-D_TG_{1T}$$

and the invariance of h_T follows from the invariance of $D_T'(D_TV_TD_T')^-D_T$, which is implied by Lemma A3.

Proof of Theorem 1.5: From the proof of Theorem 1.4, it follows that

$$(A.34)\ \sqrt{T}(\bar{b}_T - b_o) = -(H'CH)^{-1}H'C(Y_T + \alpha) + o_p(1)$$

so that

$$(A.35)\ \sqrt{T}(\bar{b}_T - \dot{b}_T) = \sqrt{T}(\bar{b}_T - b_o) - \sqrt{T}(\dot{b}_T - b_o)$$

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

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

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

where $|\ddot{b}_T - b_o| \leq |\tilde{b}_T - b_o|$ 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

$$(A.36)\ \dot{b}_T = \tilde{b}_T - (G_{2T}'C_TG_{2T})^{-1}(-G_{2T}')C_T(G_{1T} - G_{2T}\tilde{b}_T)$$

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

$$= \bar{b}_T$$

Proof of Proposition 1.6: The proof consists of showing that

(A.37) $R(Q) = R([WH \vdots L']) - q.$

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 A.5 implies

$$R(P_w'L') = R([WH \vdots L']) - q.$$

Proof of Corollary 1.7: Theorem 1.5 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 1.6,

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

Proof of Proposition 1.8: From the proof of Theorem 1.1 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

(A.38) $P_w'L'Q^-LP_w = U = V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1}.$

We know from the proof of Proposition 1.1 that $R(P_wVP_w') = r-q$. If

$R(LP_w VP_w' L') = r-q$, then by Lemma A2,

$$L'(LP_w VP_w' L')^- L = (P_w VP_w')^-$$

so that by Lemma A1

$$P_w'(P_w VP_w')^- P_w = P_w' L'(LP_w VP_w' L')LP_w.$$

Then it suffices to show

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

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

$$\text{(A.39)}\quad U - P_w'(P_w VP_w')^- P_w = F^{-1}[I - F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1} - FP_w'P_w FFP_w') - P'F]F^{-1}.$$

Now $F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1}$ and $FP_w'(P_w FFP_w')^- P_w F$ are idempotent by Lemma A.3. Further, $[FP_w'(P_w FFP_w')^- P_w F]\, F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1}] = 0$ by $P_w H = 0$, so that $I - F^{-1}H(H'F^{-1}F^{-1}H)^{-1}H'F^{-1} - FP_w'(P_w FFP_w')^- P_w F = G$ is idempotent. Therefore

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

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

Proof of Lemma 1.9: The difference $\lambda^{*2} - \lambda^2$ satisfies

(A.40) $\lambda^{*2} - \lambda^2 = \alpha' F^{-1}[FP_W'(P_W F^2 P_W')^- P_W F - FP_W'L'(LP_W F^2 P_W'L')LP_W F]F^{-1}\alpha$

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

Proof of Lemma 1.10: Note that

$$R(U_{22}) = R([0, I_k]U \begin{bmatrix} 0 \\ I_k \end{bmatrix}) = R(U \begin{bmatrix} 0 \\ I_k \end{bmatrix}).$$

Since $U = V^{-1}[I - H(H'V^{-1}H)^{-1}H'V^{-1}]$, $R(U) = r-q$, and H is a basis for N(U). Then by Lemma A5,

$$R(U \begin{bmatrix} 0 \\ I_k \end{bmatrix}) = R(\begin{bmatrix} H_1 & 0 \\ H_2 & I_k \end{bmatrix}) - q$$

$$= R(H_1) + k - q = k$$

by $R(H_1) = q$.

$L = [0\ I_k]V^{-1}$. Then by Theorem 2.1, $\lambda^2 = \alpha'P'L'(LPVP'L')^- LP\alpha$.

Note that $LPVP'L' = [0\ I_k]\ V^{-1}PVP'V^{-1} \begin{bmatrix} 0 \\ I_k \end{bmatrix} = U_{22}$ and

$LP\alpha = [0\ I_k]V^{-1}P[0, \alpha_2']' = U_{22}\alpha$ so that

$$\lambda^2 = \alpha_2'U_{22}U_{22}^- U_{22}\alpha_2' = \alpha_2'U_{22}\alpha_2.$$

Proof of Proposition 1.11: Note that $[0,I_K]V^{-1} = (V^{-1})_{22}L$, so that by non-singularikty 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 rank(Q) = rank(U_{22}) = k by Lemma 1.10. Also

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

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

Proof of Proposition 1.12: By Theorem 1.5, 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 1.6, 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, Proposition 1.6 implies

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

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 non-centrality parameter given then follows by Theorem 1.4.

Proof of Proposition 1.13: 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

(A.42) $P'L'Q^{-}LP = U[0,I_K]'B(B'[0,I_K]U[0,I_K]'B)^{-}B'[0,I_K]U$

From the proof of Proposition 1.12 the degrees of freedom of the Hausman test is $d_h^* = \text{rank}(B)$. If $d_h^* = k$, then by non-singularity of U_{22}, which implies, by Lemma A2, $B(B'U_{22}B)^{-}B' = U_{22}^{-1}$. The equation

(A.32) gives

(A.43) $P'L'Q^{-}LP = U[0,I_K]'U_{22}^{-1}[0,I_K]U$

Then asymptotic equivalence of m_T and h_T for $d_h = k$ follows from equation A.41 and A.43.

$m_T = (Y_T + \alpha)'U[0,I_k]'U_{22}^{-1}[0,I_k]U(Y_T + \alpha) + o_p(1)$ and the asymptotic equivalence of $\bar{h}_T$ and its GMM test version which implies $\bar{h}_T = (Y_T + \alpha)'U[0,I_k]'U_{22}^{-1}[0,I_k]U(Y_T + \alpha) + o_p(1)$.

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 $Y_T + \alpha = g_{1T}$. Then $\bar{h}_T = g_{1T}'U_T[0,I_k]'U_{T22}^{-1}[0,I_k]U_Tg_{1T} = m_T$.

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.42) implies

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

$$= \gamma'B'U_{22}B\gamma = \alpha_{22}'\alpha_2 = \lambda^{*2},$$

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

CHAPTER II

Maximum Likelihood Specification Testing and Instrumented Score Tests

I. Introduction

In many applications of econometrics, data is analyzed for the purpose of obtaining good estimates of a vector of parameters θ_o. In particular, it is desired that an estimator $\hat{\theta}_T$ be consistent for θ_o. If some form of model misspecification is possible, then the estimator $\hat{\theta}_T$ will often not be consistent for θ_o, so that it becomes important to have specification tests available as diagnostic tests for the validity of the economic or statistical assumptions which imply estimator consistency. Specification tests are particularly important for many econometric models which are estimated by maximum likelihood, such as limited dependent variable and non-linear simultaneous equation models, where consistency of parameter estimators depend on the validity of distributional assumptions, and where few diagnostic tests are currently available.

Specification tests for maximum likelihood estimators have been considered in Hausman (1978), Hausman and Taylor (1980), Holly (1982) and White (1982). Hausman presents a general scheme for forming specification tests based on the difference of two estimators of θ_o. White presents a specification test based on the information matrix equality, and a test based on the maximum likelihood score. All of these specification tests can be viewed as tests of specific population moment conditions. We present a moment testing framework which, for asymptotic distribution purposes, includes each of the specific specification tests presented by Hausman, Holly and White. We also derive the limiting distribution of moment specification tests for specific forms of local misspecification. We show how the asymptotic power calculations of Hausman and Taylor (1980) and Holly (1982) for maximum likelihood estimation can be extended to allow comparison of the power of any moment condition based specification tests. For example, our results allow for comparison of the local power of White's information test with the local power of a Hausman test for a particular model and form of misspecification.

An important source of moment condition tests is available in many econometric models. It is often the case that maximum likelihood estimation conditions on exogenous variables. Then the expectation of the true score (the first derivative with respect to θ of the log of the density evaluated at θ_o) of an observation, conditioned on the exogenous variables, is zero, so that functions of exogenous variables should be uncorrelated with the elements of the score vector. Moment condition tests can therefore be based on sample correlations of functions of exogenous variables and the score function, evaluated at the maximum likelihood estimator. We refer to these tests as instrumented score (IS) tests. We use our results on local power of moment condition tests to obtain sufficient conditions characterizing optimal IS tests. These optimal tests can often be interpreted as limited information versions of the classical Lagrange multiplier test.

Throughout this paper we emphasize that particular forms of moment condition tests are easy to compute. For IS tests, all that is required to compute an appropriate chi-squared statistic is the chosen instrument function of the exogenous variable for each observation, the maximum likelihood score vector for each observation and a standard regression computer package. Instrumented score tests therfore provide a simple method of obtaining diagnostic test statistics in models estimated by maximum likelihood. We illustrate the utility of IS tests by providing a menu of specification tests for the probit model.

II. Moment Specification Tests

Suppose that the econometrician observes independent observations $z_1, \ldots, z_T$, where it is assumed that the probability density function of z_t has the parametric form $f(z_t|\theta_o)$ for $t = 1, \ldots, T$. The objective of the econometrician is to estimate the kxl parameter vector θ_o. The implied normalized log-likelihood is

(2.1) $$L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ln f(z_t|\theta),$$

and the maximum likelihood estimator $\hat{\theta}_T$ is given by the solution to

(2.2) $$\max_{\theta \varepsilon \Theta} L_T(\theta)$$

where Θ is a subset of R^k. In order to perform a specification test the econometrician has available an sxl vector of functions of an observation and the parameter vector, $m(z,\theta)$, which satisfies

(2.3) $$\int m(z,\theta_o)\ f(z|\theta_o)\ d\mu(z) = 0.$$

Given a particular moment function $m(z,\theta)$, a specification test based on the sample moments $m_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} m(z_t,\theta)$ can be formed as

$$(2.4) \quad d_T = T\, m_T(\hat{\theta}_T)' Q_T^{-1} m_T(\hat{\theta}_T),$$

where Q_T is a consistent estimator of the asymptotic covariance matrix of $m_T(\hat{\theta}_T)$. The statistic d_T measures the distance of the sample moments $m_T(\hat{\theta}_T)$ from the population moment zero of equation (2.3) and should be asymptotically distributed as chi-squared with s degrees of freedom if $f(z|\theta)$ is the correct functional form for the density of z_1.

The information matrix test suggested in White (1982) is an example of a moment specification test. For the information matrix test the vector of moment functions $m(z,\theta)$ is given by

$$(2.5) \quad m_h(z,\theta) = \partial \ell nf(z|\theta)/\partial\theta_i \cdot \partial \ell nf(z|\theta)/\partial\theta_j + \partial^2 \ell nf(z|\theta)/\partial\theta_i \partial\theta_j,$$

for $h = i+j-1$, $i = 1, \ldots, j$, $j = 1, \ldots, k$. Equation (2.3) is then the equality between the gradient outer product and Hessian definitions of the Fisher information matrix. In the following sections we will discuss other choices of the moment functions $m(z,\theta)$.

In order to compare local power properties of different specification tests for a given form of misspecification, we assume that the data are not generated by the density $f(z|\theta)$. Let $h(z|\theta,\gamma)$ be a density function satisfying

(2.6) $h(z|\theta,\gamma_o) = f(z|\theta)$.

and suppose that for each T the observations $z_1,\ldots,z_T$ are independently identically distributed with probability density function $h(z|\theta_o,\gamma_T)$, where

(2.7) $\gamma_T = \gamma_o + \delta/\sqrt{T}$

The ℓx1 parameter vector γ indexes misspecification, and if $\gamma=\gamma_o$ then $f(z|\theta)$ is the correct functional form for the likelihood of each observation. For example, in the linear regression model γ may represent a vector of covariances of the linear equation residual with reduced form residuals for right-hand side variables. Equation (2.7) states that the data are generated by a sequence of local misspecification alternatives which is a convenient device for approximating power of asymptotic tests for small departures.

In order to derive the asymptotic properties of specification tests under local misspecification, we impose regularity conditions on the density $h(z|\theta,\gamma)$ and on the function $m(z,\theta)$. These regularity conditions may be of some independent interest, since they are sufficient for application of a uniform law of large numbers and a uniform central limit theorem for the maximum likelihood score due to McFadden (1980), and because the conditions also apply to models where the data is not (necessarily) continuously distributed.

For a matrix $A = [a_{ij}]$, let $|A| = \max_{i,j} |a_{ij}|$ be the supremum norm.

Assumption 2.1: The function $h(z|\theta,\gamma)$ is a measurable probability density function on a measurable space Z, and for almost all z in Z a twice continuously differentiable function of (θ,γ) in $\Theta x \Gamma$, where $\Gamma \subset R^k$ and $\Gamma \subset R^\ell$.

Assumption 2: The vector function $g(z,\theta)$ is a measurable function from Z to R^s, and for almost all z in Z a continuously differentiable function of θ in Θ.

Assumption 2.3: The sets Θ and Γ are compact, and θ_o and γ_o are elements of the interiors of Θ and Γ, respectively.

Assumptions 2.1, 2.2 and 2.3 are usual differentiability assumptions, except for the fact that the measureable space Z is allowed to have a measure other than the Lebesque measure, so that, as in Hoadley (1971), the data z need not be continuously distributed. Let $\beta' = (\theta',\gamma')$.

Assumption 2.4: There exist measurable functions $\alpha_1(z)$ and $\alpha_2(z)$ satisfying

$$|h(z|\beta)| < \alpha_1(z),\ |\ell nh(z|\beta)| < \alpha_2(z),\ |\partial\ell nh(z|\beta)/\partial\beta|^2 < \alpha_2(z)$$

$$|\partial^2\ell nh(z|\beta)/\partial\beta\partial\beta'| < \alpha_2(z),\ |m(z,\theta)|^2 < \alpha_2(z),\ |\partial m(z,\theta)/\partial\theta| < \alpha_2(z)$$

and

$$\int \alpha_1(z)d\mu(z) < +\infty \quad \text{and} \int \alpha_2(z)\alpha_1(z)d\mu(z) < +\infty$$

Also, the set of z in Z such that $h(z|\beta) > 0$ is independent of β.

Let $s(z,\theta) = \partial\ell nf(z|\theta)/\partial\theta$ be the score vector for an observation, for the density $f(z|\theta) = h(z|\theta,\gamma_o)$ on Z. Define the following matrices:

$$J = E_{\theta_o}[s(z,\theta_o)s(z,\theta_o)'],$$

$$K = E_{\theta_o}[s(z,\theta_o)\partial\ell nh(z|\theta_o,\gamma_o)/\partial\gamma'],$$

$$C = E_{\theta_o}[m(z,\theta_o)s(z,\theta_o)'],$$

$$N = E_{\theta_o}[m(z,\theta_o)\partial\ell nh(z|\theta_o,\gamma_o)/\partial\gamma'],$$

$$M = E_{\theta_o}[m(z,\theta_o)m(z,\theta_o)'],$$

$$V = \begin{bmatrix} J & C' \\ C & M \end{bmatrix},$$

$Q = M - CJ^{-1}C'$ and

$U = N - CJ^{-1}K.$

Note that the matrix J is the Fisher information matrix.

Assumption 2.5: The matrix V is non-singular.

Assumption 2.6: If $\theta \neq \theta_o$, then $A = \{z : f(z|\theta) \neq f(z|\theta_o)\}$ satisfies

$$\int_A f(z|\theta_o)d\mu(z) > 0.$$

Assumption 2.6 is a global identification assumption. The next assumption specifies the data generating process.

Assumption 2.7: For each T the observations $z_1, \ldots, z_T$ are independent with z_t having density $h(z_t|\theta_o, \gamma_T)$, $t = 1, \ldots, T$, for $\gamma_T = \gamma_o + \delta/\sqrt{T}$.

For notational convenience we have suppressed an extra T subscript on z_t.

Assumption 2.8: The sequence of matrices Q_T satisfies plim $Q_T = Q$.

Theorem 2.1: If Assumptions 2.1-2.8 are satisfied then d_T converges in distribution to a non-central chi-squared distribution with s degrees of freedom and noncentrality parameter

$$(2.8) \qquad \nu = \delta'U'Q^{-1}U\delta$$

In order to compute argument specification test statistic in practice a consistent estimator Q_T of the asymptotic covariance matrix Q is required A simple Q_T is available. Let

$$(2.9) \qquad R = \begin{bmatrix} s(z_1,\hat{\theta}_T)' \\ \vdots \\ s(z_T,\hat{\theta}_T)' \end{bmatrix}, \quad F = \begin{bmatrix} m(z_1,\hat{\theta}_T)' \\ \vdots \\ m(z_T,\hat{\theta}_T)' \end{bmatrix}.$$

Let I_T be an TxT identity matrix.

Proposition 2.2: If Assumptions 2.1-2.7 are satisfied then

$$Q_T = \frac{1}{T} F'[I_T - R(R'R)^{-1}R'] F = \frac{1}{T} F'F - \frac{1}{T} F'R(R'R)^{-1}R'F$$

satisfies assumption 2.8.

Note that Q_T of Proposition 2.2 is positive semi-definite and is the difference of two terms. Further, Q_T involves only the moment function $m(z,\theta)$ and the score $s(z,\theta)$, so that derivatives of $m(z,\theta)$ and $s(z,\theta)$ are not needed for computing d_T. For example use of Q_T of proposition 2.2 in forming White's information matrix test means that third derivatives of the log of the density need not be computed. This simplification is achieved by using an outer product estimator of J and an outer product estimator of $E_{\theta_o}[\partial m(z,\theta_o)/\partial\theta]$. Whenever the functions $m(z,\theta)$ satisfy the identity of assumption 2.2,

(2.10) $\int m(z,\theta)f(z|\theta)d\mu(z) \equiv 0,$

and the regularity conditions of assumptions 2.1-2.4, differentiation of both sides of equation (2.10) yields

(2.11) $E_{\theta_o}[\partial m(z,\theta_o)/\partial\theta] = - E_{\theta_o}[m(z,\theta_o)s(z,\theta_o)']$.

This equality is exactly analogous to the equality between the outer product and Hessian definitions of the Fisher information matrix.

It is shown in Chapter 1 that a general one-step theorem yields the equivalence of Hausman and moment tests under local misspecification. To indicate how our results can be applied to Hausman tests, we briefly present a formal framework for maximum likelihood Hausman testing which complements that of White (1982). Suppose that $s \geq k$ and that an estimator $\tilde{\theta}_T$ satisfies

(2.12) $\sqrt{T}\, D_T m_T(\tilde{\theta}_T) = o_p(1),$

where plim $D_T = D$. The estimator $\tilde{\theta}_T$ is a generalized method of moments estimator, as recently discussed in Hansen (1982), and for asymptotic distribution purposes most econometric estimators can be viewed as being of this form.

Proposition 2.3. If assumptions 2.1-2.8 are satisfied, and in addition DC is non-singular and plim $\tilde{\theta}_T = \theta_o$, then

(2.13) $\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) = \sqrt{T}\ (DC)^{-1} Dm_T(\hat{\theta}_T) + o_p(1)$.

Proposition 2.3 means that a Hausman test statistic h_T given by

(2.14) $h_T = T(\hat{\theta}_T - \tilde{\theta}_T)' S_T^{-1} (\hat{\theta}_T - \tilde{\theta}_T)$,

where plim $S_T = S = (DC)^{-1} DMD'(C'D')^{-1} - J^{-1}$ is the asymptotic covariance matrix of $\sqrt{T}\ q_T = \sqrt{T}\ (\hat{\theta}_T - \tilde{\theta}_T)$, is asymptotically equivalent under local misspecification to a moments test with moment functions

(2.15) $\bar{m}(z,\theta) = Dm(z,\theta)$

Application of Theorem 2.1 to the moment function $\bar{m}(z,\theta)$ of equation (2.15) gives the distribution of h_T, under local misspecification, in terms of the parameters of the underlying data generating process. Further, by proposition 2.3 a moment specification test statistic which is asymptotically equivalent to the Hausman test based on q_T is given by

$$(2.15)\quad d_T^h = Te'FD_T'[D_TF'FD_T' - D_TF'R(R'R)^{-1}R'FD_T']^{-1}D_TF'e/e'e$$

$$= TR_h^2,$$

where R_h^2 is the R-squared from a regression of e on (R,FD_T').

It is also of some independent interest to note that there is a consistent estimator of S which is simple and positive semi-definite. Let

$$(2.17)\quad \tilde{F} = \begin{bmatrix} m(z_1,\tilde{\theta}_T)' \\ \vdots \\ m(z_T,\tilde{\theta}_T)' \end{bmatrix}.$$

Then a consistent estimator of S is given by

$$(2.18)\quad S_T = (\tfrac{1}{T}\, D_T\tilde{F}'R)^{-1} D_T\tilde{F}'\tilde{F}D_T'(R'\tilde{F}D_T')^{-1} - (\tfrac{1}{T}\, R'R)^{-1}$$

where S_T is positive semi-definite, the difference of two terms and consists of averages of outer products.

An important property of moment specification tests is that they may not be consistent. It is shown in Chapter 1 that a moment specification test fails to reject with probability approaching one on an ℓ-s dimensional submanifold of R^ℓ, when $\ell > s$. Therefore if the dimension of the misspecification parameter space is large enough any particular moment specification test, or Hausman test, will not be consistent. Further, the dimension of the set on which a moment test fails has dimension equal to the dimension of the misspecification space minus the degrees of freedom of the test. The set of failure points of the specification test is of smaller dimension than ℓ, so that the set of failure points has Lebesque measure zero in R^{ℓ}. Since the set of failure points has measure zero, the issue of consistency may not be important in applications, although the power of moment specification test can be quite low for a large set of points in the misspecification parameter space. It is important in applications to consider the power properties of specification tests when picking a particular test.

III. Instrumented Score Tests

Many econometric models have more structure than that presented in the last section. Frequently, the component variables of an observation z are partitioned as

(2.19) $z = (y,x)$,

where y consists of endogenous and x consists of exogenous variables. We formalize the notion of exogenous variables by specifying that the density factors into two pieces,

(2.20) $f(z|\theta) = f_1(y|x,\theta)f_2(x)$,

where $f_1(y|x,\theta)$ can be thought of as the conditional density of y given x, and $f_2(x)$ as the marginal density of x. In terms of the terminology of Engle, Hendry and Richard (1983), the i.i.d. observations and equation (2.20) imply x is strictly exogenous. The function $f_1(y|x,\theta)$ will be assumed to be known, while $f_2(x)$ remains unknown. Note that $f_2(x)$ does not depend on θ, so that the normalized log-likelihood can be taken to be

(2.21) $L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ln f(y_t|x_t,\theta)$,

and the maximum likelihood estimator $\hat{\theta}_T$ solves

$$(2.22)\quad \max_{\theta \text{ in } \Theta} L_T(\theta)$$

This type of factorization is common in econometrics. For example, in the discrete choice framework of Manski and McFadden (1981), the conditional probabilities of a finite set of alternatives given exogenous variables and parameters is assumed known, while the distribution of exogenous variables remains unspecified.

To justify our interpretation of $f_1(y|x,\theta)$ as a conditional density, we add the following assumption to those of section two.

Assumption 2.9: The measure space Z is the product of two measure spaces Y and X. The functions of $f_1(y|x,\theta)$ and $f_2(x)$ are measurable functions of z, and for almost all z, $f_1(y|x,\theta)$ is a twice continuously differentiable function of θ in Θ. Further, with probability one (w.p.1), $1 = \int f_1(y|x,\theta)d\mu(y)$ for all θ in Θ.

Formally, what we will mean by "conditional on x" is the following. Let C be the sub σ-algebra consisting of the product of the σ-algebra for X and the trivial algebra for Y. Then when the density is given by $f(z|\bar{\theta})$ for $\bar{\theta}$ in Θ, and $g(z)$ is an integrable random variable, for θ in Θ define the function $E_{\bar{\theta}}(g(z)|x)$ by

(2.23) $E_{\bar{\theta}}(g(z)|x) = E_{\bar{\theta}}[g(z)|C]$

(see Chung (1974, chp. 9)). Then the interpretation of $f_1(y|x,\theta)$ as a conditional density is justified by the following result.

Lemma 2.4: If $E_{\bar{\theta}}[g(z)] = \int g(z)f_1(y|x,\bar{\theta})f_2(x)d\mu(z) < +\infty$ and assumption 2.9 is satisfied, then w.p.1 $E_{\bar{\theta}}[g(z)|x] = \int g(z)f_1(y|x,\bar{\theta})d\mu(y)$.

The importance of the factorization of equation (2.20) for maximum likelihood specification testing stems from the following simple result.

Lemma 2.5: If assumption 2.1, 2.4 and 2.9 are satisfied, then for each $\bar{\theta}$ in the interior of Θ,

(2.24) $E_{\bar{\theta}}[s(z,\bar{\theta})|x] = 0$ w.p.1

Lemma 2.5 implies that when the score function is evaluated at θ_o, it has expectation zero, conditional on the exogenous variables. Therefore the true score $s(z,\theta_o)$ should be uncorrelated with functions of the exogenous variables. This fact can be used to form moment condition functions of the form

$$(2.25)\quad m(z,\theta) = w(x,\theta)s(z,\theta),$$

where $w(x,\theta)$ is a sxk matrix, which can be used to form specification tests, as described in section two. The moment function $m(z,\theta)$ consists of sums of cross-products of functions of the exogenous variables and elements of the score vector, so that it is natural to refer to maximum likelihood specification tests using moment conditions of the form given in equation (2.25) as instrumented score (IS) tests. Frequently in econometrics there is a kxn vector $u(x,\theta)$ and a nxl vector $r(z,\theta)$ such that the elements of the score vector satisfy

(2.26) $u(x,\theta)r(z,\theta) = s(z,\theta)$ where for all $\bar{\theta}$ in the interior of Θ

$$(2.27)\quad E_{\bar{\theta}}\left[r(z,\bar{\theta})|x\right] = 0.$$

For example, in the probit model, where y is either zero or one and x is a lxk vector of real-valued variables, and where $\text{Prob}_\theta(y=1|x) = \Phi(x\theta)$ for $\Phi(\varepsilon)$ the c.d.f. of the standard normal distribution, we have

$$(2.28)\quad \ell n f_1(y|x,\theta) = y\ell n\Phi(x\theta) + (1-y)\ell n[1-\Phi(x\theta)].$$

Let $\lambda(\varepsilon) = \phi(\varepsilon)/[\Phi(\varepsilon)(1-\Phi(\varepsilon))]$, where $\phi(\varepsilon)$ is the p.d.f. of a standard normal distribution. Then differentiation of equation (2.28) yields

$$(2.29)\quad s(z,\theta) = \lambda(x\theta)x'[y-\Phi(x\theta)].$$

Since

(2.30) $E_\theta[y-\Phi(x\theta)|x] = 0$

for the probit model we can take $u(x,\theta) = \lambda(x\theta)x'$ and $r(z,\theta) = y - \Phi(x\theta)$. It will convenient in what follows to consider the more general set of functions $r(z,\theta)$ rather than just $s(z,\theta)$. The more general type of moment condition function we will consider is

(2.31) $m(z,\theta) = w(x,\theta)r(z,\theta)$.

We will also refer to specification tests based on moment functions of this form as IS tests. To guarantee that $m(z,\theta)$ given in equation (2.31) satisfies the assumptions of section two, we impose the following assumption.

<u>Assumption 2.10</u>: The sum of functions $w(x,\theta)$ and the nx1 vector of functions $r(z,\theta)$ are measurable functions of z and for almost all z in Z are continuously differentiable functions of θ in Θ. Also there exists measurable functions $\alpha_1(z)$ and $\alpha_3(z)$ such that $|w(x,\theta)|^4 \leq \alpha_3(z)$, $|r(z,\theta)|^4 \leq \alpha_3(z)$, $|\partial r(z,\theta)/\partial\theta|^2 \leq \alpha_3(z)$, $|\partial w_{ij}(x,\theta)/\partial\theta| \leq \alpha_3(z)$ for $i = i,\ldots,s$, $j=1,\ldots,m$, $|\ell nh(z)\beta| \leq \alpha_3(z)$, $|\partial\ell nh(z|\beta)/\partial\beta|^2 \leq \alpha_3(z)$,
$|\partial^2\ell nh(z|\beta)/\partial\beta\partial\beta'| \leq \alpha_3(z)$, and $h(z|\beta) \leq \alpha_1(z)$ for all β in $\Theta \times \Gamma$, and $\int\alpha_1(z)d\mu(z) < +\infty$ and $\int a_3(z)\alpha_1(z)d\mu(z) < +\infty$. Further

(2.32) $E_{\bar{\theta}}\,[r(z,\bar{\theta})\,|\,x] = 0$

w.p.1 for all $\bar{\theta}$ in Θ.

<u>Lemma 2.6</u>: If assumption 2.10 is satisfied, then $m(z,\theta) = w(x,\theta)r(z,\theta)$ and $h(z|\beta)$ satisfy assumptions 2.2 and 2.4.

The use of IS specification tests does not depend on the form of misspecifications. Subject to the regularity conditions we have imposed moment functions can always be formed using the score function (or the more fundamental function $r(z,\theta)$) and the exogenous variables. However, considering particularly the potential failure of moment tests, it seems wise to consider carefully the power properties of an IS test in a particular application. We can obtain a sufficient characterization of the optimal IS test for a particular form of misspecification.

For the optimal IS test we require $s \geq \ell$ in order to obtain consistency of the moment specification test, as discussed in section two. The asymptotic power of a chi-squared test is decreasing in the degrees of freedom, so that the optimal consistent test should have degrees of freedom ℓ. The asymptotic power is also increasing in the non-centrality parameter. A sufficient characterization of the ℓxn instrument functions which maximize the non-centrality parameter is given in the following result. Let x_1 be a vector of exogenous variables which are a subset of those included in x.

Theorem 3.7: Suppose that for $r(z,\theta)$ and the pair of matrix functions $w(x_1,\theta)$ and $\bar{w}(x_1,\theta)$, 2.7, 2.9 and 2.10 are satisfied, and in addition the $\ell \times m$ matrix of functions, $\bar{w}(x_1,\theta)$ satisfies

$$(2.33)\quad E_{\theta_o}[r(z,\theta_o)r(z,\theta_o)'|x_1]\bar{w}(x_1,\theta_o)' = E_{\theta_o}[r(z,\theta_o)\partial \ell nh(z|\theta_o,\gamma_o)/\partial\gamma'|x_1]$$

and

$$(2.34)\quad E_{\theta_o}[s(z,\theta_o)r(z,\theta_o)'|x_1]\,\bar{w}(x_1,\theta_o)' = E_{\theta_o}[s(z,\theta_o)\partial \ell nh(z|\theta_o,\gamma_o)/\partial\gamma'|x_1].$$

Then for any δ in R^{ℓ}, the non-centrality parameter for the moment condition specification test based on $m(z,\theta) = \bar{w}(x_1,\theta)r(z,\theta)$ is at least as large as the non-centrality parameter for the moment specification test based on $w(x_1,\theta)r(z,\theta)$.

Note that theorem 2.7 gives the form of the optimal instrument functions of the subset of exogenous variables x_1. Some of the exogenous variables may not be observed by the econometrician, so that x_1 is relevant for formation of IS tests. In this case we will have $f_1(y|x,\theta) = f_1(y|x_1,\theta)$.

To interpret the form of the optimal instrument functions, note that if there exists $\bar{w}(x,\theta)$ such that

$$(2.35)\quad \bar{m}(z,\theta_o) = r(z,\theta_o)\bar{w}(x,\theta_o) = \partial \ell nh(z|\theta_o,\gamma_o)/\partial\gamma$$

then equations (2.33) and (2.34) are satisfied and the moment function $\bar{m}(z,\theta)$ can be taken to be $\partial \ell nh(z|\theta,\gamma_o)/\partial\gamma$. The IS test based on $\bar{m}(z|\theta)$ will be an LM test of H_o: $\gamma = \gamma_o$. More generally, all the information needed to form $\partial \ell nh(z|\theta,\gamma_o)/\partial\gamma$ may not be available (see the next section for examples). For notational simplicity, suppose $\ell = 1$ and $r(z,\theta) = s(z,\theta)$. Then when the kxl vector $\bar{w}(x,\theta)$ satisfies equation (2.33) we have

$$(2.36)\quad E[\, s(z,\theta_o)(\partial \ell nh(z|\theta_o,\gamma_o)/\partial\gamma - s(z,\theta_o)\bar{w}(x,\theta_o))\,|x] = 0$$

Therefore, conditional on x, $\bar{w}(x,\theta_o)s(z,\theta_o)$ is the best linear predictor of $\partial \ell nh(z|\theta_o,\gamma_o)/\partial\gamma$.

Theorem 2.7 gives a method of obtaining optimal instrument functions $\bar{w}(x,\theta)$ for particular forms of misspecification. To obtain the instrument functions $\bar{w}(x,\theta)$, the conditional expectations given in equations (2.33) and (2.34) have to be computed, and these conditional expectations involve $\partial \ell nh(z|\theta_o,\gamma_o)/\partial\gamma$. Evidently, computation of these optimal instrument functions requires complete specification of the form of possible misspecification. In applications, the form of possible misspecification is usually not known. One way of picking the instrument functions is to choose optimal instruments for misspecification which is known, through theoretical or empirical experience, to cause large bias. For example, in the truncated normal regression model Hurd (1979) has shown that moderate amounts of heteroskedasticity can cause large bias, so that an optimal test for heteroskedasticity may be appropriate. However, in many applications, several types of possible specification error may be of concern. The specification errors of concern can be incorporated into the IS test statistic by stacking all of the orthogonality condition functions into one vector of orthogonality condition functions, and using this one larger dimensional vector to form the IS test statistic. The results of chapter I imply that if any one of the individual types of misspecification occur by itself, the non-centrality parameter for the joint test will be the same as the non-centrality parameter for the optimal test of the individual type of misspecification, and local power loss will thus be entirely due to the larger degrees of freedom. However, this power loss can be great and robustness of the test against many forms of misspecification must be balanced with power against particular forms of misspecification.

The results of this section also provide for an optimal generalization of the Hausman test suggested by White (1980) which is based on the difference of the ordinary least squares and a weighted least squares estimator. In the linear model with normally distributed disturbances the score vector is given by,

$$(2.37)\quad s(z,\theta) = \begin{bmatrix} x'(y-x\beta)/\sigma^2 \\ [(y-x\beta)^2-\sigma^2]/2\sigma^4 \end{bmatrix},$$

where $\theta = (\beta,\sigma^2)$ and, conditional on x,y is distributed normally with mean $x\beta$ and variance σ^2. A weighted least squares estimator of β is a method of moments estimator with moment functions given by

$$(2.38)\quad m(z,\theta) = v(x)^2 x'(y-x\beta) = [v(x)^2 I,0]s(z,\theta),$$

where $v(x)$ is the weighting function. The WLS versus OLS Hausman test is therefore an instrumented score test, by proposition 2.3.

It is of some independent interest to note that equation (2.38) implies that the WLS estimator is an instrumental variables estimator with instruments $v(x)^2x'$, so that following Hausman (1978), the Hausman test statistic can be computed as a test statistic for the inclusion of $v(x)^2x'$ in a regression of y on $(x',v(x)^2x')'$. In particular, as in Chapter 1, a version of the Hausman test statistic can be computed as T times the R-squared from a regression of the OLS residuals on $(x',v(x)^2x')$.

IV. Specification Tests for Probit.

To illustrate the results of previous sections we now present a discusion of specification testing in the context of binary probit estimation. Our results are also of substantial independent interest, since a range of diagnostic tests is provided, all of which are extremely easy to implement, and binary probit is a commonly used non-linear maximum likelihood procedure.

To set up the probit model, let y^* be an unobserved latent varible, and suppose that the data are given by

$$(2.39)\quad z = (y,x),\ x = (x_1,x_2),\ x_1 = (x_{11},\ x_{12}),$$

where y is an indicator variable which is one if $y^*>0$ and zero othewise, x_1 is a 1 x k vector of variables which are observed by the econometrician and x_2 is a 1 x p vector of unobserved variables. Let the distribution of y^* conditional on x and θ be normal with mean $x_1\theta$ and variance 1. Then

$$(2.40)\quad f(y|x,\theta) = \Phi(x_1\theta)^y[1-\Phi(x_1\theta)]^{1-y}.$$

The score function is

$$(2.41)\quad s(z,\theta) = \lambda(x_1\theta)x'[y-\Phi(x_1\theta)],$$

where $\Phi(\varepsilon)$ and $\lambda(\varepsilon)$ have been previously defined. A fundamental orthogonality condition function in the independent probit model is

(2.42) $r(z,\theta) = y - \Phi(x_1\theta)$,

which follows by noting that the expectation of y given x and θ is $\Phi(x_1\theta)$.

Let $\hat{\theta}_T$ be the MLE for the probit model. We will assume that assumptions 2.1, 2.3, 2.5, 2.7, 2.9 and 2.10 are satisfied in the context of the probit model. IS tests for probit use orthogonality condition functions of the form

(2.43) $m(z,\theta) = w(x_1\theta)[y-\Phi(x_1\theta)]$

From equation (2.43) and the form of the score function, given in equation (2.41) it is apparent that any IS test is identical with a LM test for the inclusion of extra variables

(2.44) $$c_T(z_t) = \frac{\Phi(x_{1t}\hat{\theta}_T)[1-\Phi(x_{1t}\hat{\theta}_T)]}{\phi(x_{1t}\hat{\theta}_T)} w(x_{1t},\hat{\theta}_T)$$

That an asymptotically equivalent test can be computed as a likelihood ratio (LR) test is the result of the following proposition.

<u>Proposition 2.8:</u> If assumptions 2.1, 2.3, 2.5, 2.7, 2.9 and 2.10 are

satisfied, the likelihood ratio test for inclusion of $c_T(z_t)$ in the probit model with exogenous variables x_{1t} and $c_T(z_t)$ is asymptotically equivalent to the IS test for the $m(z,\theta)$ of equation (2.43).

Proposition 2.8 follows by the same arguments which lead to the asymptotic equivalence of the LM and LR tests for a parametric hypothesis. An IS test for probit can be obtained by one additional probit maximization using $c_T(z_t)$ as extra variables and the computation of the likelihood ratio test statistic.

We consider five possible sources of misspecification and optimal tests for them. We assume throughout this discussion that x_1 includes a constant. A parameterization of heteroskedasticity can be obtained by assuming that conditional on x, θ and γ y^* is normally distributed with mean zero and variance $\sigma(x,\gamma)^2$, where $\sigma(x,0) = 1$ w.p.1. A parameterization of non-normality as presented in Ruud (1981), can be obtained by assuming that conditional on x, $\varepsilon = y^*-x_1\theta$ has c.d.f. $\Phi(\gamma+\varepsilon+\gamma_2\varepsilon^2+\gamma_3\varepsilon^3)$, where γ_2 and γ_3 satisfy $1+2\gamma_2\varepsilon+3\gamma_3{}^2>0$ for all ε. A parameterization of a simultaneity or errors in variables problem can be obtained by assuming that

(2.45) $x'_{11} = x_{12}\pi_1+x_2\pi_2+v,$

where, conditional on x_{12}, x_2, θ, (π_1,π_2), γ and $\sum$, (2.46) $\binom{\varepsilon}{v} \sim N([\begin{smallmatrix}0\\0\end{smallmatrix}], [\begin{smallmatrix}1 & \gamma'\\ \gamma & \Sigma\end{smallmatrix}])$.

A parameterization of sampling contamination can be obtained by assuming that the sample is choice-based with probability of drawing y=1 equal to, for $Q = \int \Phi(x_1\theta_o)f_2(x)d\mu(x)$,

(2.47) $H = \gamma + \int \Phi(x_1\theta_o)f_2(x)d\mu(x)$.

Finally, an unobserved omitted variables problem can be parameterized by assuming that, conditional on x , θ and γ , y* is normally distributed with mean $x_1\theta+x_2\gamma$ and variance 1.

The kernel of the log-likelihood of an observation under misspecification, the forms of the optimal instrument function as given in theorem 2.7 and the form of the extra variables to be included as in proposition 2.8 are as follows.

Heteroskedasticity:

$\ell nh(z|\beta) = y\, \ell n\Phi(x_1\theta/\sigma(x,\gamma)) + (1-y)\ell n[1-\Phi(x_1\theta/\sigma(x,\gamma))]$

$w(x,\theta) = \lambda(x_1\theta)\cdot x_1\theta\cdot E[\partial\sigma(x,0)\partial\gamma\,|x_1]$

$c_T(z_t) = x_{1t}\hat{\theta}_T \cdot E[\partial\sigma(x_t,0)/\partial\gamma\,|x_{1t}]'$

Non-Normality:

$$\ell nh(z|\beta) = y\ell n\Phi(x_1\theta + \gamma_2(x_1\theta)^2 + \gamma_3(x_1\theta)^3) + (1-y)\ell n[1-\Phi(x_1\theta) + \gamma_2(x_1\theta)^2 + \gamma_3(x_1\theta)^3)]$$

$$\bar{w}(x,\theta) = \lambda(x_1\theta) \cdot [(x_1\theta)^2 , (x_1\theta)^3]'$$

$$c_T(z_t) = [(x_{1t}\hat{\theta}_T)^2 , (x_{1t}\hat{\theta}_T)^3].$$

Simultaneity:

$\ell nh(z|\beta)$: Omitted: see, e.g., Ruud (1981)

$$\bar{w}(x,\theta) = \lambda(x_1\theta)[x_{11}' - x_{12}\pi_1 - E(x_2|x_1)\pi_2]$$

$$c_T(z_t) = [x_{11}' - x_{12}'\pi_1 - E(x_2|x_1)\pi_2].$$

Sampling Contamination (Choice-based sample):

$$h(z|\beta) = y\{\ell n\Phi(x_1\theta) + \ell n(1+\gamma/Q)\} + (1-y)\{\ell n[1-\Phi(x_1\theta)] + \ell n[1-\gamma/(1-Q)]\}$$

$$\bar{w}(x,\theta) = 1$$

$$c_T(z_t) = 1/\lambda(x_{1t}\hat{\theta}_T).$$

Unobserved Omitted Variable:

$h(z|\beta) = y\ell n\Phi(x_1\theta + x_2\gamma) + (1-y)\ell n[1-\Phi(x_1\theta + x_2\gamma)]$

$\overline{w}(x,\theta) = \lambda(x_1\theta)E(x_2|x_1)'$

$C_T(z_t) = E(x_{2t}|x_{1t})'$.

To form the optimal test for heteroskedasticity, $E[\partial\sigma(x,0)/\partial\gamma\,|x_1]$ is required. Let $\tilde{x}$ be the non constant components of x_1. If $\sigma(x,\gamma) = h(q(\tilde{x}_1)\gamma)$ where $h(0) = 1$ and h is continuously differentiable and $q(\tilde{x}_1)$ a $1\times\ell$ vector of functions of $\tilde{x}_1$, then

(2.48) $E[\partial\sigma(x,0)/\partial\gamma\,|x_1] = \partial\sigma(x,0)/\partial\gamma = q(\tilde{x}_1)'$

A general method of obtaining a good test against heteroskedasticity will be to take

(2.49) $w(x,\theta) = x_1\theta \cdot q(\tilde{x}_1)$

where $q(\tilde{x}_1)$ is an $\ell x1$ vector consisting of the elements of $\tilde{x}_1$ (and possibly cross-products) of the elements of $\tilde{x}_1$. This procedure amounts to a linear (or quadratic) approximation of $E[\partial\sigma(x,0)/\partial\gamma\,|x_1]$, which holds exactly if $\sigma(x,\gamma)$ has the functional form mentioned earlier.

For the test of non-normality, the implied density for ε when $\gamma_2 \neq 0$ and $\gamma_3 \neq 0$ is

$$(2.50)\quad f(\varepsilon) = \phi(\gamma_0 + \varepsilon + \gamma_2\varepsilon^2 + \gamma_3\varepsilon^3) \cdot (1+2\gamma_2\varepsilon + 3\gamma_3\varepsilon^2),$$

so that the distribution of ε has fatter tails than the normal, and if $\gamma_2 \neq 0$ is non-symmetric. It is also of interest to note that this specification test for non-normality is asymptotically equivalent to an LM test against the two parameter Pearson family of distributiaons, as is apparent from comparison of the orthogonality condition function

$$(2.51)\quad m(z,\theta) = \lambda(x_1\ \theta) \cdot [y-\Phi(x_1\theta)]\ [(x_1\theta)^2, (x_1\theta)^3],$$

with the formula for the derivatives of the probit likelihood with respect to the Pearson parameters, presented in Lee (1980).

A joint test of non-normality and heteroskedasticity can be obtained by setting

$$(2.52)\quad c_T(z_t) = ((x_{1t}\hat{\theta}_T) \cdot \tilde{x}_{1t},\ (x_{1t}\hat{\theta}_T)^3),$$

where the $(x_{1t}\hat{\theta}_T)^2$ is omitted due linear dependence, or

$$(2.53)\quad c_T(z_t) = = (x_{1t}\hat{\theta}_T) \cdot (\tilde{x}_{1t},\ v(\tilde{x}_{1t})),$$

where $v(\tilde{x}_{1t})$ consists of the distinct elements of $\tilde{x}_{1t}'\tilde{x}_{1t}$. As discussed in the previous section, these joint tests will have the same non-centrality parameter against just non-nomality test, but will have lower power than the non-normality test due to the larger degrees of freedom. It is also interesting that the IS test with $c_T(z_t)$ as in equation (2.53) is the information matrix test of White (1982), so that in the context of the probit model the information matrix test is a joint test for normality and heteroskedasticity.

In the formulas for the test against simultaneity, (Π_1, Π_2) is the matrix of coefficients of the least squares regressions of the elements of the contaminated variables x_{11t} on x_{12t} and the predicted instruments $E(x_{2t}|x_{1t})$. If some elements of x_{2t} are observed, then these variables should be used directly as instruments. Otherwise $E(x_{2t}|x_{1t})$ should be approximated as closely as possible, by non-linear functions of x_{1t}. If auxiliary sample information is available, so that for some observations on the entire vector x_t are available these observations could be used to estimate the function $E(x_{2t}|x_{1t})$. Auxiliary sample information could also be used to obtain the optimal set of test variables for the unobserved omitted variable test given above.

The test for sample selection contamination has the following interesting interpretation. The sample average of the orthogonality condition function used in forming this specification test is given by

$$(2.54)\ \frac{1}{T}\sum_{t=1}^{T} m(z_t,\hat{\theta}_T) = (T_1/T) - \sum_{t=1}^{T} \Phi(x_{1t}\hat{\theta}_T)/T,$$

where T_1 is the number of observations with $y_t = 1$ so that this test is based on the difference of the sample proportion of observations with $y_t = 1$ and the maximum likelihood predicted proportion. More generally, a test of sampling contamination can be based on orthogonality condition functions of the form $\frac{1}{T}\sum_{t=1}^{T}[y_t - \Phi(x_{1t}\hat{\theta}_T)]q(x_{1t})$ is correlated with the sampling process. A complete treatment of tests of sampling contamination will be developed in future research.

Finally, some remarks on specification testing of the censored normal model are appropriate. Suppose instead of only observing

$$\tilde{y} = \begin{cases} y^* & \text{if } y^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

is also observed. Then maximum likelihood estimation of $\theta = (\beta,\sigma)'$ using the likelihood function

$$(2.54)\quad L_T(\theta) = L_T(\gamma,\sigma) = \frac{1}{T}\sum_{t=1}^{T}\{y_t[-1/2\ln\sigma^2 - 1/2\sigma^2(\tilde{y}_t - x_{1t}\beta)^2]$$
$$+(1-y_t)\ln[1-\Phi(x_{1t}\beta/\sigma)]$$

is appropriate, as discussed in Amemiya (1973). Specification tests can still be based on the $r(z,\theta)$ function

$$(2.55)\quad r(z,\theta) = y - \Phi(x_1\beta/\sigma)$$

although the conditional moments of $\tilde{y}$ can be used to obtain more powerful tests. It can be shown that the optimality arguments for the tests of sample contamination and heteroskedasticity based on $r(z,\theta)$ still hold. Nelson (1981) has considered a specification test for the normal censored regression model based on a $y-\Phi(x_1\delta/\sigma)$. In the framework of our paper, Nelson's test is an optimal test for sampling contamination when $q(x_1)$ is taken to be x_1'.

A more appropriate test for misspecification resulting from heteroskedasticity should be of the form of the test we have presented. We leave to future research a comparison of relative power of the tests we have presented under various forms of misspecification.

Appendix to Chapter Two

We first give two uniform convergence results, due to McFadden, which are useful in proving convergence under a sequence of local alternatives. The first result is a doubly uniform weak law of large numbers. For a function $g(z,\theta)$, a density $h(z|\beta)$ and a sequence of observations $\underset{\sim}{z} = (z_1, z_2, \ldots)$ define

$$f_T(\theta) = f_T(\underset{\sim}{z},\theta) = \frac{1}{T}\sum_{t=1}^{T} g(z_t,\theta)$$

and where expectations exist

$$\phi(\theta,\beta) = E(g(z,\theta)|\beta) = \int g(z,\theta)h(z|\beta)d\mu(z)$$

$$V(\theta,\beta) = E(g(z,\theta)g(z,\theta)\ |\beta) - \phi(\theta,\beta)\phi(\theta,\beta)'$$

Lemma A1: Suppose $g(z,\theta)$ is a measureable function on a measureable space Z, and for almost all z in Z a continuous function of θ in Θ, a compact subset of R^q. Suppose $(h|\beta)$ is a measureable probability density on Z and for almost all z in Z a continuous function of β in B, a subset of R^q. Suppose $\alpha(z) = \sup|g(z,\theta)|$ and $\gamma(z) = \sup h(z|\beta)$ satisfy

(A.1) $\int \gamma(z)d\mu(z) < +\infty, \quad \int \alpha(z)\gamma(z)d\mu(z) < +\infty$

Suppose z is a sequence of independent observations from the probability $h(z|\beta)$. Then $f_T(\theta)$ converges in probability to $\phi(\theta,\beta)$ uniformly on $\Theta x B$, i.e., for $\varepsilon,\delta > 0$ there exists $T(\varepsilon,\delta)$ such that if $n > n(\varepsilon,\delta)$, for all β in B

(A.2) $\quad \text{Prob}(\sup|f_T(\theta) - \phi(\theta,\beta)| > \varepsilon|\beta) < \delta$

Further, $\phi(\theta,\beta)$ is continuous on $\Theta x B$.

The second result is a uniform central limit theorem.

Lemma A2: Suppose $g(z,\theta)$ is a measureable function on a measureable space Z, and for almost all z in Z, a continuous function of θ in Θ. Suppose $h(z|\beta)$ is a measureable function of z and for almost all z a continuous function of β in B. Suppose $\alpha(z) = \sup|g(z,\theta)|$ and $\gamma(z) = \sup h(z|\beta)$ satisfy

(A.3) $\quad \int \gamma(z)d\mu(z) < \infty, \int \alpha(z)^2\gamma(z)d\mu(z) < \infty$

Suppose z is a sequence of independent observations from the probability $h(z|\beta)$. Suppose $V(\theta,\beta)$ in uniformly non-singular (White (1982)) on $\Theta x B$. Then

$\sqrt{T}V(\theta,\beta)^{-1/2}(f_T(\theta)-\phi(\theta,\beta))$ converges in distribution to $Y \sim N(0,I)$ uniformly on $\Theta x B$; i.e., given $\varepsilon > 0$, there exists $T(\varepsilon)$ such that $T > T(\varepsilon)$ implies

(A.4) $$\sup_{\theta x B}\sup_{y}|\text{Prob}(\sqrt{T}V(\theta,\beta)^{-1/2}(f_T(\theta) - \phi(\theta,\beta)) \leq y) - \text{Prob}(Y \leq y)| < \varepsilon.$$

Proof of Theorem 2.1: We first prove that plim $\hat{\theta}_T = \theta_o$. (Note that probability statements depend on sample size due to the sequence of local alternatives, so that plim $\hat{\theta}_T = \theta_o$ means $\lim_{T\to\infty} \text{Prob}_T(|\hat{\theta}_T - \theta_o| > \varepsilon) = 0$ for each $\varepsilon > 0$.) Assumptions 2.1 and 2.3 specify that $L_T(\theta)$ is measureable in the data and continuous in θ and that Θ is compact, so that Lemma 2 of Jennrich (1969) implies that a measureable θ_T exists. Define $Q(\theta) = \int \ln f(z|\theta)f(z|\theta_o)d\mu(z)$, where the integral exists by assumption 2.4, which implies $\int |\ln f(z|\theta)f(z|\theta_o)|d\mu(z) \leq \int \alpha_2(z)\alpha_1(z)d\mu(z) < +\infty$. Then assumption 2.6 implies that $Q(\theta)$ has a unique maximum at θ_o, by the classical information inequality (see Rao 1973, p. 59).

By assumptions 2.1 and 2.4 $\ln f(z|\theta)$ and $h(z|\theta_o,\gamma)$ are measureable in z and continuous in θ and γ respectively, and

(A.5) $\int \sup_{\gamma} h(z|\theta_o,\gamma)d\mu(z) < +\infty$, $\int \sup_{\theta}|\ln f(z|\theta)| \sup_{\gamma} h(z|\theta_o,\gamma)d\mu(z) < +\infty$.

Define $\kappa(\theta,\gamma) = \int \ln f(z|\theta)h(z|\theta_o,\gamma)d\mu(z)$. Then by Lemma A1, for any $\varepsilon,\delta > 0$ there exists $T(\varepsilon,\delta)$ such that if $T > T(\varepsilon,\delta)$

(A.6) $\text{Prob}(\sup_{\theta}|L_T(\theta) - \kappa(\theta,\gamma)| > \varepsilon|\gamma) < \delta$

for any γ in Γ. In particular, for any $\varepsilon > 0$

(A.7) $\lim_{T\to\infty} \mathrm{Prob}(\sup_{\theta}|L_T(\theta) - \kappa(\theta,\gamma_T)| > \varepsilon\,|\gamma_T) = 0$

Further, by Lemma A1 $\kappa(\theta,\gamma)$ is continuous on $\Theta\times\Gamma$. By the compactness hypothesis of assumption 2.3, $\kappa(\theta,\gamma)$ is uniformly continuous, from which it follows that

(A.8) $\lim_{T\to\infty} \sup_{\theta} |\kappa(\theta,\gamma_T) - \kappa(\theta,\gamma_o)| = 0.$

Since $Q(\theta) = \kappa(\theta,\gamma_o)$ and

(A.9) $\sup_{\theta}|L_T(\theta) - Q(\theta)| \leq \sup_{\theta} |L_T(\theta) - \kappa(\theta,\gamma_T)| + \sup_{\theta} |\kappa(\theta,\gamma_T) -$

$Q(\theta)|$

equations (A.7) and (A.8) imply

(A.10) $\mathrm{plim} \sup_{\theta}|L_T(\theta) - Q(\theta)| = 0$

Then by a convergence in probability version of Lemma 3 of Amemiya (1973) (see McFadden, 1980, p. 20), $\mathrm{plim}\ \hat{\theta}_T = \theta_o$.

We finish the proof of Theorem 2.1 by verifying that assumptions 2.1-2.8 are sufficient for assumptions 1.1-1.16 of chapter 1. Let $s(z,\theta) = \partial \ln f(z|\theta)/\partial\theta$ and $g(z,\theta)' = (s(z,\theta)', m(z,\theta)')$.

In terms of the notation of chapter 1, let

$b = \theta$, $b_o = \theta_o$, $c = \gamma$, $c_o = \gamma_o$, $B = \Theta$, $g(z,b) = g(z,\theta)$, $W_T = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$,

and $L_T = [0, I_s]$. Then $h(b,c) = \int g(z,\theta)h(z|\theta_o,\gamma)d\mu(z)$.

Consider the equation

(A.11) $\sqrt{T}\partial L_T(\hat{\theta}_T)/\partial\theta = o_p(1)$

By plim $\hat{\theta}_T = \theta_o$ and assumption 2.3, which specifies that θ_o is the interior of Θ, $\hat{\theta}_T$ satisfies the first order conditions $\partial L_T(\hat{\theta}_T)/\partial\theta = 0$ with probability approaching one. Thus, $\hat{\theta}_T$ satisfies (A.11), and in terms of the notation of chapter 1 we can take $\hat{b}_T = \hat{\theta}_T$.

Let $A = \{z{:}h(z|\beta) > 0\}$. By assumption 2.4, A does not depend on β. On A, $[\partial \ell nf(z|\theta)/\partial\theta]\cdot f(z|\theta) = \partial f(z|\theta)/\partial\theta$. Also,

(A.12) $$\sup_\theta |[\partial \ell nf(z|\theta)/\partial\theta]\cdot f(z|\theta)| \leq [\sup_\theta|\partial \ell nf(z|\theta)/\partial\theta|]\cdot \sup_\theta f(z|\theta) \leq \alpha_2(z)\alpha_1(z)$$

so that by integrability of $\alpha_2(z)\cdot\alpha_1(z)$ differentiating the identity $1 = \int f(z|\theta)d\mu(z)$ under the integral sign is allowed, so that

(A.13) $\int s(z,\bar{\theta})f(z|\bar{\theta})d\mu(z) = 0.$

for all $\bar{\theta}$ in the interior of Θ. Evaluation of equation (A.13) at $\bar{\theta} = \theta_o$ and assumption 2.2 imply

(A.14) $h(b_o,c_o) = \int g(z,\theta_o)f(z|\theta_o)d\mu(z) = 0$

so that assmption 1.1 of chapter 1 is satisfied. Assumption 1.2 of chapter 1 follows directly from assumptions 2.1, 2.2 and 2.4.

To see that assumption 1.3 of chapter 1 is satisfied, note that, by the definition of $|x|$

(A.15) $\sup_\theta |\partial m(z,\theta)/\partial\theta \cdot h(z|\theta_o,\gamma)| \leq \sup_\theta |\partial s(z,\theta)/\partial\theta| \cdot \sup_\gamma h(z|\theta_o,\gamma)$

$+ \sup_\theta |\partial m(z,\theta)/\partial\theta|\} \sup_\gamma h(z|\theta_o,\gamma) \leq 2\alpha_2(z)\alpha_1(z)$

so that h(b,c) has a vector of partial derivatives

(A.16) $\partial h(b,c)/\partial b = \partial/\partial\theta[\int g(z,\theta)h(z|\theta_o,\gamma)d\mu(z)]$

$= \int[\partial g(z,\theta)/\partial\theta]h(z|\theta_o,\gamma)d\mu(z)$

$= E[\partial g(z,b)/\partial b]$

where we drop the θ_o subscript on the expectation for notational conveience. Further, since assumption 2.4 implies $\mu(Z/A) = 0$, $\partial h(z|\theta_o\gamma)/\partial\gamma = \partial \ell n h(z|\theta_o,\gamma)/\partial\gamma \cdot \quad h(z|\theta_o,\gamma)$. Also,

(A.17) $\sup_\gamma |g(z,\theta)\partial h(z|\theta_o,\gamma)/\partial\gamma'|$

$\leq \sup_\gamma |g(z,\theta)\partial \ell n h(z|\theta_o,\gamma)/\partial\gamma'| \sup_\gamma h(z|\theta_o,\gamma)$

$\leq \{\sup_\gamma |s(z,\theta)\partial \ell n h(z|\theta_o,\gamma)/\partial\gamma'|$

$+ \sup_\gamma |m(z,\theta)\partial \ell n h(z|\theta_o,\gamma)/\partial\gamma'|\} \ \alpha_1(z)$

$\leq 2\alpha_2(z)\alpha_1(z)$

where the third inequality follows from $|xy| \leq 1/2|x|^2 + 1/2|y|^2$ and the implication of assumption 2.4 that $|s(z,\theta)|^2$, $|m(z,\theta)|^2$ and $|\partial \ell n h(z|\theta_o,\gamma)/\partial\gamma|^2$ are each dominated by $\alpha_2(z)$. Therefore h(b,c) has a vector of partial derivatives

(A.18) $\partial h(b,c)/\partial c = \int g(z,\theta)\ \partial \ln h(z|\theta_o,\gamma)/\partial\gamma' h(z|\theta_o,\gamma)d\mu(z)$

By the dominance conditions of equations (A.15) and (A.17) and the continuity conditions of assumptions 2.1 and 2.2, the dominated convergence theorem implies that $\partial h(b,c)/\partial b$ and $\partial h(b,c)/\partial c$ are continuous in (b,c) so that h(b,c) is differentiable. Now, equation (A.15) and assumption 2.4 imply

(A.19) $\int \sup_{\gamma} h(z|\theta_o,\gamma)d\mu(z) < +\infty$,

$\int \sup_{\theta}|\partial g(z,\theta)/\partial\theta| \sup_{\gamma} h(z|\theta_o,\gamma)d\mu(z) < +\infty$

so that by assumptions 2.1 and 2.2 the hypotheses of Lemma A1 are satisfied and $\partial g_T(b)/\partial b = \frac{1}{T}\sum_{t=1}^{T}\partial g(z,\theta)/\partial\theta$ converges in probability uniformly in θ and γ to $M(\theta,\gamma) = \int \partial m(z,\theta)/\partial\theta \cdot h(z|\theta_o,\gamma)\ d\mu(z)$, and $M(\theta,\gamma)$ is continuous in θ and γ. Then

(A.20) $\text{plim} \sup_{\theta} |\frac{1}{T}\sum_{t=1}^{T}\partial m(z_t,\theta)/\partial\theta - M(\theta,\gamma_T)| = 0$

and, since uniform continuity of $M(\theta,\gamma)$ on the compact set $\Theta x\Gamma$ implies

(A.21) $\lim \sup_{\theta}|M(\theta,\gamma_o) - M(\theta,\gamma_T)| = 0$

it follows that

$$(\text{A.22})\quad \operatorname{plim} \sup_{\theta} \left| \frac{1}{T} \sum_{t=1}^{T} \partial m(z_t,\theta)/\partial\theta - M(\theta,\gamma_0) \right| = 0$$

In terms of the notation of chapter 1, H = $\int \partial g(z,\theta_0)/\partial\theta\, f(z|\theta_0)d\mu(z)$ and $H'WH = \{\int[\partial^2 \ell n f(z|\theta)/\partial\theta\partial\theta']f(z|\theta_0)d\mu(z)\}^2$, where $W = \operatorname{plim} W_T = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$ is positive semi-definite. Further, by assumption 2.4

$$\begin{aligned}(\text{A.23})\quad |\partial/\partial\theta[s(z,\theta)f(z|\theta)]| &= |s(z,\theta)s(z,\theta)' \\ &\quad + \partial^2 \ell n f(z|\theta)/\partial\theta\partial\theta'|f(z|\theta) \\ &\leq [|s(z,\theta)s(z,\theta)'| \\ &\quad + |\partial^2 \ell n f(z|\theta)/\partial\theta\partial\theta'|]f(z|\theta) \leq 2\alpha_2(z)\alpha_1(z).\end{aligned}$$

By the integrability of $\alpha_2(z)\cdot\alpha_1(z)$ specified by assumption 2.4, equation (A.23) implies that differentiation under the integral is allowed when differentiating both sides of the identity (A.13) to obtain

$$(\text{A.24})\quad 0 = \int s(z,\bar\theta)s(z,\bar\theta)'f(z|\bar\theta)d\mu(z) + \int \partial^2 \ell n f(z|\bar\theta)/\partial\theta\partial\theta' f(z|\bar\theta)d\mu(z)$$

for all $\bar\theta$ in the interior of Θ. Equation (A.24), evaluated at $\bar\theta = \theta_0$, implies that $H'WH = J^2$, so that $H'WH$ is non-singular by assumption 2.5. Finally, noting that $\operatorname{plim} L_T = [0, I_s]$, we see that assumption 1.4 of chapter is satisfied.

We now use Lemma A2 to verify that assumption 1.5 of chapter 1 is satisfied. Note that

(A.25) $|g(z,\theta_o)g(z,\theta_o)'h(z|\theta_o,\gamma)| \leq \alpha_2(z)\alpha_1(z)$, $|g(z,\theta_o)h(z|\theta_o,\gamma)|$
$\leq (1+\alpha_2(z))\alpha_1(z)$

so that by the integrability of $\alpha_2(z)\alpha_1(z)$ and $\alpha_1(z)$ and the dominated convergence theorem $V(\gamma)$ is continuous in γ, where

(A.26) $V(\gamma) = \int g(z,\theta_o)g(z,\theta_o)'h(z|\theta_o,\gamma)d\mu(z) - h(b_o,c)h(b_o,c)'$

In terms of the notation of Newey (1982)

(A.27) $V = \lim_{T\to\infty} TE[g_T(b_o)g_T(b_o)' - h(b_o,c_T)h(b_o,c_T)'] = \lim_{T\to\infty} V(\gamma_T)$

$= V(\gamma_o)$

where the second equality follows from independence of z_t and z_s for $t\neq s$. The matrix V is non-singular by assumption 2.5. By $V(\gamma)$ continuous in γ and $V(\gamma_o)$ non-singular $V(\gamma)$ is uniformly non-singular in a neighborhood U of γ_o. Also, assumption 2.4 implies

(A.28) $h(z|\theta_o,\gamma) \leq \alpha_1(z)$, $|g(z,\theta_o)| \leq \alpha_2(z)^{1/2}$

Then since $\alpha_1(z)$ and $\alpha_2(z)\alpha_1(z)$ are integrable, and by the continuity hypotheses of assumptions 2.1 and 2.2 the hypotheses of Lemma A2 are satisfied for $\beta=\gamma$,

$h(z|\beta) = h(z|\theta_o,\gamma)$, $g(z) = m(z,\theta_o)$, $\gamma(z) = \alpha_2(z)^{1/2}$, B=U and $V(\beta) =$

$V(\gamma)$. Since

$\lim_{T\to\infty} \gamma_n = \gamma_T$, γ_T lies in U for large enough n, so that the conclusion of Lemma A2 implies

$$(A.29)\ \sqrt{n}\ V(\gamma_T)^{-1/2}\Big[\frac{1}{T}\sum_{t=1}^{T} m(z_t,\theta_o) - \int m(z,\theta_o)h(z|\theta_o,\gamma_T)d\mu(z)\Big] \overset{d}{\to} Y$$

where $Y \sim N(0,I)$. Since $\lim_{T\to\infty} V(\gamma_T) = V(\gamma_o)$, pre multiplying equation (A.29) by $V(\gamma_T)^{1/2}$ yields

$$(A.30)\ \sqrt{n}\Big[\frac{1}{T}\sum_{t=1}^{T} m(z_t,\theta_o) - \int m(z,\theta_o)h(z|\theta_o,\gamma_T)d\mu(z) \overset{d}{\to} V(\gamma_o)^{1/2}\ Y$$

where $V(\gamma_o)^{1/2}Y \sim N(0,V(\gamma_o))$.

Finally, we consider assumption 1.6 of chapter 1. From assumption 2.2,

$$(A.31)\ \int m(z,\bar{\theta})f(z|\bar{\theta})d\mu(z) = 0$$

for all $\bar{\theta}$ in the interior of Θ. Also,

$$(A.32)\ |\partial/\partial\theta[m(z,\theta)f(z|\theta)]| = |[\partial m(z,\theta)/\partial\theta + m(z,\theta)s(z,\theta)']f(z|\theta)| \leq \{|\partial m(z,\theta)/\partial\theta| + |m(z,\theta)s(z,\theta)|\} \times f(z|\theta) \leq \partial\alpha_2(z)\alpha_1(z)$$

so that by integrability of $\alpha_2(z)\cdot\alpha_1(z)$ differentiation under the integral of both sides of the identity (A.31) is allowed, and this differentiation yields

$$(A.33)\ 0 = \int \partial m(z,\bar{\theta})/\partial\theta\, f(z|\bar{\theta})d\mu(z) + \int m(z,\bar{\theta})s(z,\bar{\theta})'f(z|\bar{\theta})d\mu(z)$$

for all $\bar{\theta}$ in the interior of Θ. Evaluating equations (A.33) and (A.24) at $\bar{\theta} = \theta_o$ gives, in terms of the notation of chapter 1.

$$(A.34)\ H = \int \partial g(z,\theta_o)/\partial\theta\, f(z|\theta_o)d\mu(z) = -[J,C']'$$

and, therefore

$$(A.35)\ LP_W = L - LH(H'WH)^{-1}H'W = [-CJ^{-1},\ I_s]$$

so that

$$(A.36)\ Q = LP_WVP_W'L' = M - CJ^{-1}C'$$

is non-singular, and there is no choice involved in picking a generalized inverse, since plim $Q_T = Q$ implies plim $Q_T^{-1} = Q^{-1}$.

To finish the proof of Theorem 2.1, note that evaluating equation (A.18) at $\theta = \theta_o$, $\gamma = \gamma_o$ yields

$$(A.37)\ \partial h(b_o,c_o)/\partial c = \begin{bmatrix} K \\ N \end{bmatrix}$$

so that

$$(A.38)\ LP_W[\partial h(b_o,c_o)/\partial c]\delta = U\delta$$

and the conclusion follows from Theorem 2.1 of chapter 1.

Proof of Proposition 2.2: As in the proof that $\text{plim} \sup_{\theta} |L_T(\theta)-Q(\theta)| = 0$, it follows from assumption 2.1-2.4 that

$$(A.39) \quad \text{plim} \sup_{\theta} \Big| \frac{1}{T} \sum_{t=1}^{T} g(z_t,\theta)g(z_t,\theta)' - \int g(z,\theta)g(z,\theta)'f(z|\theta_o)d\mu(z)\Big| = 0$$

Further, assumptions 2.1-2.4 implies that $\int g(z,\theta)g(z,\theta)'f(z|\theta_o)d\mu(z)$ is continuous in θ, by continuity of $g(z,\theta)$ and the dominated convergence theorem. Then by plim $\theta_T = \theta_o$ and a convergence in probability version of Lemma 4 of Amemiya (1973),

$$(A.40) \quad \text{plim} \frac{1}{T} \sum_{t=1}^{T} g(z_t,\hat{\theta}_T)g(z_t,\hat{\theta}_T)' = V$$

Consistency of Q_T for V then follows by the non-singularity of J.

Proof of Proposition 2.3: We again use the results of chapter 1.

Since $\tilde{\theta}_T$ satisfies equation (2.12), we can show as in the poof of theorem 2.1 of chapter 1 that

$$(A.41) \quad \sqrt{T}(\tilde{\theta}_T - \theta_o) = -(DC)^{-1}C\sqrt{T}m_T(\theta_o) + o_p(1)$$

Now apply Theorem 1.5 of chapter 1, and the conclusion follows.

Proof of Lemma 2.4: Let $q(x) = \int g(z)f_1(y|x,\bar{\theta})d\mu(y)$. By the Fubini Theorem $q(x)$ is X measureable, implying $q(x)$ is C measureable. By the Fubini theorem and assumption 2.9

$$(A.42)\quad E_{\bar{\theta}}[q(x)] = \int q(x)f_2(x)d\mu(x) = E_{\bar{\theta}}[g(z)] < +\infty$$

so that $q(x) < +\infty$ w.p.1. For any set A in C, either $A = \phi xB$ or $A=YxB$ for some measureable subset, B, of $\bar{X}$. For $A=\phi xB$, $1_A = 0$ for all z in Z (where 1_A is the indicator for the set A) and

$$(A.43)\quad E_{\bar{\theta}}[1_A \cdot q(x)] = 0 = E[1_A \cdot g(z)]$$

For $A=YxB$, $1_A = 1_B$, and by the Fubini theorem and assumption 3.1

$$(A.44)\quad E_{\bar{\theta}}[1_A \cdot q(x)] = \int 1_B \cdot q(x)f_2(x)d\mu(x)$$

$$= \int [\int g(z)f_1(y|x,\bar{\theta})d\mu(y)]1_B \cdot f_2(x)d\mu(x)$$

$$= E_{\bar{\theta}}[1_A \cdot g(z)]$$

Proof of Lemma 2.5: Let $D = \{x:f_2(x) \geq 0\}$. Then $Prob(D) = 1$. Further, for x in D, $A_x = \{y|f_1(y|x,\theta) > 0\}$ is independent of θ by assumption 2.1. On A_x, $\partial/\partial\theta[f_1(y|x,\theta)] = s(z,\theta)f_1(y|x,\theta)$. Note that

$$(A.45)\quad |s(z,\theta)|\ f_1(y|x,\theta) \leq (\alpha_2(z) + 1)\frac{\alpha_1(z)}{f_2(x)}$$

By the Fubini theorem, integrability of $\alpha_1(z)$ and $\alpha_2(z)\cdot\alpha_1(z)$ imply that

(A.46) $\int \alpha_1(z)d\mu(y) < \infty, \quad \int \alpha_2(z)\alpha_1(z)d\mu(y) < +\infty$

for almost all x. Let $\bar{D}$ be a subset of D such that equations (A.46) and $1 = \int f_1(y|x,\theta)d\mu(y)$ for all θ in Θ hold, and such that Prob $(\bar{D})$ = 1. Then for x in $\bar{D}$ equations (A.45) and (A.46) allow differentiation under the integral sign to obtain, for $\bar{\theta}$ in the interior of Θ,

(A.47) $0 = \int_{A_x} s(z,\bar{\theta})f_1(y|x,\bar{\theta})d\mu(y) = E_{\bar{\theta}}[s(z,\bar{\theta})|x]$

<u>Proof of Lemma 2.6</u>: Note that

(A.48) $|w(x,\theta)r(z,\theta)| \leqslant \frac{n}{2}|w(x,\theta)|^2 + \frac{n}{2}|r(z,\theta)|^2 \leqslant n(\alpha_3(z) + 1)$

so that for any $\bar{\theta}$ and θ both in Θ, $E_{\bar{\theta}}[|w(x,\theta)r(z,\theta)|] < +\infty$. For any $\bar{\theta}$ in Θ Chung (1974) theorems 9.1.3 and 9.1.5, and $E_{\bar{\theta}}[r(z,\theta)|x] = 0$

(A.49) $E_{\bar{\theta}}[w(x,\bar{\theta})r(z,\bar{\theta})] = E_{\bar{\theta}}[E_{\bar{\theta}}[w(x,\bar{\theta})r(z|\bar{\theta})|x]]$

$$= E_{\bar{\theta}}[w(x,\bar{\theta})E_{\bar{\theta}}[r(z,\bar{\theta})|x]] = 0.$$

Also, $w(x,\theta)r(z,\theta)$ is continuously differentiable, and the dominance conditions of assumption 2.4 follow from

$$\text{(A.50)}\quad |\partial/\partial\theta[w(x,\theta)r(z,\theta)]| \leq |w(x,\theta)\partial r(z,\theta)/\partial\theta| + |\sum_i \partial w_i(z,\theta)/\partial\theta \cdot r_i(z,\theta)| \leq n\,\alpha_3(z) + n(\alpha_3(z) + 1)$$

and

$$\text{(A.51)}\quad |w(x,\theta)r(z,\theta)|^2 \leq \frac{n}{2}\,|r(z,\theta)|^4 \leq n\alpha_3(z)$$

and the dominance hypotheses of assumption 2.10 on $h(z|\beta)$.

Proof of Theorem 2.7: Let $v = \delta'U'Q^{-1}U\delta$ be the non-centrality parameter for the moments test with $m(z,\theta) = w(x,\theta)r(z,\theta)$ and $\overline{v} = \delta'\overline{U}'\overline{Q}^{-1}\overline{U}\delta$ be the non-centrality parameter for the moments test with $m(z,\theta) = \overline{w}(x,\theta)r(z,\theta)$, where $U, V, \overline{Q}$ and $\overline{Q}$ are defined as in section two for the respective moment condition functions. Define the $\ell + s + k$ vector.

$$\text{(A.52)}\quad R(z) = \begin{bmatrix} R_1(z) \\ R_2(z) \\ R_3(z) \end{bmatrix} = \begin{bmatrix} \overline{w}(x,\theta_o)r(z,\theta_o) \\ w(x,\theta_o)r(z,\theta_o) \\ s(z,\theta_o) \end{bmatrix}$$

Then $P = E[R(z)R(z)']$ exists by assumption 2.10 and is clearly positive semi-definite. We will drop the θ_o subscript on E_{θ_o} for convenience. By Chung (1974) Theorems 9.1.3 and 9.1.5

(A.53) $Q_{11} = E[\overline{w}(x,\theta_o)r(z,\theta_o)r(z,\theta_o)'\overline{w}(x,\theta_o)']$

$= E[\overline{w}(x,\theta_o)E[r(z,\theta_o)r(z,\theta_o)'|x]w(x,\theta_o)']$

$= E[E[\partial\ell nh(z|\theta_o,\gamma_o)/\partial\gamma \cdot r(z,\theta_o)'|x]\ \overline{w}(x,\theta_o)']$

$= E[\partial\ell nh(z|\theta_o,\gamma_o)/\partial\gamma \cdot r(z,\theta_o)'\overline{w}(x,\theta_o)']$

(A.54) $Q_{12} = E[\partial\ell nh(z|\theta_o,\gamma_o)/\partial\gamma \cdot r(z,\theta_o)'w(x,\theta_o)']$

(A.55) $Q_{13} = E[\partial\ell nh(z|\theta_o,\gamma_o)/\partial\gamma \cdot s(z,\theta_o)]$

It follows that

(A.56) $\overline{U} = \overline{V} = Q_{11}^{-1} - Q_{13}Q_{33}Q_{13}'$

and

(A.57) $U = P_{12}' - P_{23}P_{33}^{-1}\ P_{13}'$, $Q = P_{22} - P_{23}P_{33}^{-1}\ P_{23}'$

Straightforward, but tedious, algebra then yields

(A.58) $\overline{U}'\overline{V}^{-1}\overline{U} - U'V^{-1}U = \overline{V} - U'V^{-1}U = PQP'$

where

(A.59) $G = [I_\ell,\ - U'Q^{-1}]\ x \begin{vmatrix} I_\ell & 0 & -P_{13}P_{33}^{-1} \\ 0 & I_s & -P_{23}P_{33}^{-1} \end{vmatrix}$

Therefore, since P is positive semi-definite,

$$(A.60)\quad \delta'\overline{U}'\overline{Q}^{-1}\overline{U}\delta - \delta'U'Q^{-1}U\delta = \delta'GPG'\delta > 0$$

CHAPTER III

Identification and Estimation of Simultaneous Equations with Covariance Restrictions

1. Introduction

The problems of identification and estimation of a simultaneous equation system with restrictions on the disturbance covariance matrix have a relatively long history in econometrics. The original Cowles foundation research left the problem of identification with covariance restrictions unresolved. Further work on this question was done by Fisher (1966), Rothenberg (1971), and Wegge (1965). Recently, Hausman and Taylor (1982) have provided necessary and sufficient conditions for identification in the single equation (LIML) case, as well as examination of some important examples of identification of an entire system of equations. Maximum likelihood estimation of a system of simultaneous equations with normally distributed disturbances (FIML) when the disturbance covariance matrix is diagonal was considered in the original Cowles Foundation research, Koopmans, Rubin and Leipnik (1950, pp. 154-211) and by Malinvaud (1970, pp. 678-682) and Rothenberg (1973, pp. 77-79 and pp. 94-115). Hausman and Taylor (1982) have obtained an interpretation of FIML with covariance restrictions as an instrumental variables estimator. In addition to the predetermined variables, the instruments for each equation include the residuals corresponding to the disturbances of other equations which are specified to be uncorrelated with the disturbance of the equation.

In this chapter we discuss identification and estimation of a system of simultaneous equations with covariance restrictions in the context of generalized method of moments (GMM) estimation. The estimation method we consider is a natural generalization of instrumental variable estimators to the covariance restrictions case. Instrumental variable estimators are GMM estimators where the moment functions which have expectation zero consist of cross-products of instruments and disturbances. When a pair of disturbances are specified to be uncorrelated, the product of these disturbances can also be used as a moment function which has expectation zero. These moment functions can also be interpreted as instrumental variable type moment functions, where for $\sigma_{ij} = 0$, the disturbance for equation i qualifies as an instrument for equation j and the disturbance for equation j qualifies as an instrument for equation i.

Identification of the parameters of a system of simultaneous equations can also be considered in the context of GMM estimation. We derive a general rank condition for local identification of a vector of parameters given a set of moment condition functions. When applied to a simultaneous equation system with covariance restrictions this rank condition gives necessary conditions which are interpretable in terms of an assignment of residuals as instruments to equations, where one assignment of a residual to an equation is made for each covariance restriction. This assignment condition is formulated by Hausman and Taylor (1982). We argue that an assignment condition is also sufficient for local identification. When the disturbances are normally distributed our rank condition is equivalent to non-singularity of the information matrix, so that local identification given the set of moment functions consisting of cross products of instruments and residuals with residuals is equivalent to the usual definition of local identification.

When we turn to estimation we find that the optimal GMM estimator, given by Hansen (1982), for a simultaneous equation system is asymptotically equivalent to FIML when the disturbances are normally distributed. The optimal GMM estimator will be more efficient than three-stage least-squares (3SLS), when the 3SLS estimator exists, even if the disturbances are not normally distributed. Unfortunately, GMM estimators which use the covariance restrictions are obtained as the solution to a set of non-linear equations, so that it is useful to have two-step procedures available. We present an optimal instrumental variables estimator, augmented three-stage least-squares (A3SLS), which uses estimated residuals as instruments, and show that in practice it is likely that A3SLS will be as efficient as the optimal GMM estimator, and thus will attain the Cramer-Rao lower bound asymptotically, if the disturbances are normally distributed.

Finally, it seems desirable to be able to have powerful testing procedures to test over-identifying covariance restrictions. We apply the results of Chapter 1 to obtain optimal tests of over-identifying covariance restrictions.

II. Identification

The standard linear simultaneous equations model where all identities have been substituted out can be written

(3.1) $YB_o + Z\Gamma_o = U,$

where Y is the TxM matrix of jointly endogenous variables, Z is the TxK matrix of predetermined variables and U is the TxM matrix of structural disturbances of the system. The model has T observations, M equations and K predetermined variables. We assume that B_o is non-singular, plim $(Z'U/T) = 0$, plim $(Z'Z/T) = C$, and for $X = (Y,Z)$, plim $(X'X/T)$ is non-singular. Let $(Y_t, Z_t, U_t) = (X_t, U_t)$ be a row of $(Y,Z,U) = (X,U)$. We assume that the rows of U are independently and identically distributed and that $E(U_t'U_t) = \Sigma$ is non-singular. We will allow for restrictions on the elements of Σ which are of the form $\sigma_{ij} = 0$ for $i \neq j$.

It is convenient to consider the simultaneous equation system in regression form. We normalize each equation so that $B_{ii} = 1$, $i=1,\ldots,M$. Also we impose exclusion restrictions on the coefficients of each equation and rewrite the system (3.1) as

$$(3.2) \qquad y_i = X_i\delta_i + u_i \quad ; \quad i=1,\ldots,M,$$

$$X_i = [Y_i, Z_i] \quad , \; \delta_i' = (\beta_i', \gamma_i').$$

For each i, Y_i is a $T\times r_i$ matrix of included right hand side endogenous variables, Z_i is a $T\times s_i$ matrix of included predetermined variables and δ_i is partitioned conformably with X_i. The number of right hand side variables in each equation will be denoted by $q_i = r_i + s_i$ and the number of elements of $\delta' = (\delta_1', \ldots, \delta_M')$ by $q = \sum_{i=1}^{M} q_i$. Throughout we will stick to the case of exclusion restrictions, although extension of our results to the case of linear restrictions is as straightforward as usual.

We will also assume that L restrictions of the form $\sigma_{ij} = 0$ for some $i \neq j$ are satisfied, where $1 \leq L \leq M(M-1)/2$. Let S be a $L\times M^2$ selection matrix which is defined as follows. For two $M\times 1$ vectors a and b, $S\cdot(a \otimes b)$ consists of L elements $a_i b_j$ where $i < j$ and $\sigma_{ij} = 0$. Let $u_t' = (u_{1t}, \ldots, u_{Mt})'$. The fundamental vector of moment functions for the simultaneous equations system with covariance restrictions is given by the $MK + L$ vector

$$(3.3) \qquad g(X_t, \delta) = \begin{bmatrix} u_t' \otimes Z_t' \\ S\cdot(u_t' \otimes u_t') \end{bmatrix}.$$

Let the $MK + L$ vector of functions $g(\delta)$ be defined by

(3.4) $g(\delta) = \text{plim}\ \frac{1}{T} \sum_{t=1}^{T} g(X_t, \delta)$,

where the probability limit exists by previous assumptions. For δ equal to the true parameter vector δ_o,

(3.5) $g(\delta_o) = 0$.

A natural criterion for identification of δ_o in this setting is that the set of equations $g(\delta) = 0$ has a unique solution at δ_o. As expected, this identifiability criterion is the same as the usual criteria for the linear simultaneous equations model. Let the reduced form be given by

(3.6) $Y = Z(-\Gamma_o B_o^{-1}) + UB_o^{-1} = Z\Pi + V$

and let $A_o = (B_o', \Gamma_o')'$ and $A = (B', \Gamma')'$ where B and Γ are obtained from the regression equation (3.2). Then straightforward calculation gives

(3.7) $\text{plim}\ \frac{1}{T} \sum_{t=1}^{T} u_t' \otimes Z_t' = \text{vec}\{C[\Pi, I](A - A_o)\}$,

and for $\sigma_{ij} = 0$

(3.8) $\text{plim}\ \frac{1}{T} \sum_{t=1}^{T} u_{ti} u_{tj} = (A_i - A_{oi})'[\Pi, I]'C[\Pi, I](A_j - A_{oj})$

$+ (B_i - B_{oi})'\Omega(B_j - B_{oj})$,

where A_i and A_{oi} are the ith columns of A and A_o respectively, and Ω =

$\text{plim}(V'V/T) = (B_o')^{-1}\Sigma B_o^{-1}$. Then by the non-singularity of C, $g(\delta) = 0$ if and only if

(3.9) $[\Pi, I](A-A_o) = 0$

and for all $\sigma_{ij} = 0$

(3.10) $(B_i - B_{oi})'\Omega(B_j - B_{oj}) = 0$

Therefore, the criterion that the moment equations $g(\delta) = 0$ have a unique solution at the true parameter δ_o is an identifiability criterion which is equivalent to the usual criterion: see Hausman and Taylor (1982).

<u>Definition 3.1</u>: The parameter vector δ_o is said to be identified with respect to a particular vector of moment condition functions $g(X_t, \delta)$ if the equations $g(\delta) = 0$ have a unique solution equal to δ_o.

It is intuitively clear that if δ_o is identified with respect to a vector of moment condition functions then δ_o is identified. Let θ be a vector of parameters which, along with δ, determine the data-generating process (DGP). We assume that $\text{plim}\ \frac{1}{T}\sum_{t=1}^{T} g(X_t, \delta) = g(\delta; \tilde{\delta}, \tilde{\theta})$ exists for all δ when the parameters of the DGP are given by $(\tilde{\delta}, \tilde{\theta})$ and that $g(\tilde{\delta}; \tilde{\delta}, \tilde{\theta}) = 0$.

Proposition 3.1: If δ_o is identified with respect to $g(X_t,\delta)$, then δ_o is identified.

Proof: When the parameters of the DGP process are (δ_o,θ_o), if δ_o is not identified then there exists $(\tilde{\delta},\tilde{\theta})$ with $\tilde{\delta} \neq \delta_o$ such that the DGP implied by $(\tilde{\delta},\tilde{\theta})$ has the same distribution as the DGP for (δ_o,θ_o). In particular, for all T, $\varepsilon > 0$,

$$\text{Prob}\left(\left|\frac{1}{T}\sum_{t=1}^{T} g(X_t,\tilde{\delta})\right| \geq \varepsilon;\tilde{\delta},\tilde{\theta}\right) = \text{Prob}\left(\left|\frac{1}{T}\sum_{t=1}^{T} g(X_t,\tilde{\delta})\right| \geq \varepsilon;\delta_o,\theta_o\right) \text{ so that}$$

$$0 = g(\tilde{\delta};\tilde{\delta},\tilde{\theta}) = g(\tilde{\delta};\delta_o,\theta_o) = g(\tilde{\delta}).\quad \text{QED.}$$

Of course identification of δ_o does not imply that δ_o is identified with respect to a particular set of moment functions. For example, in the linear simultaneous equations model where the disturbances have a non-normal distribution, the reduced form parameters (Π,Ω) may not determine completely the distribution of X, so that the parameters (A,Σ) may be identified even though they are not identified from equations (3.9) and (3.10).

As usual, the non-linearity of $g(\delta)$ means that rank conditions arising from local identification considerations may be much easier to verify than global identification.

Definition 3.2: The parameter vector δ_o is said to be locally identified with respect to a vector of moment condition functions if there is a neighborhood N of δ_o such that the equation $g(\delta) = 0$ has a unique solution in N at δ_o.

Suppose that the functions $g(\delta)$ are continuously differentiable in δ and $\partial g/\partial\delta(\delta)$ has constant rank in a neighborhood of δ_o. When these regularity conditions are satisfied, the following rank condition is equivalent to local identification. Let $G = \partial g/\partial\delta(\delta_o)$.

Proposition 3.2. If $g(\delta)$ is continuously differentiable and $\partial g/\partial\delta(\delta)$ has constant rank in a neighborhood of δ_o, then δ_o is locally identified with respect to $g(X_t,\delta)$ if and only if rank $(G) = q$.

Proof: The proof follows immediately from Theorem 5.A.1 of Fisher (1966). QED

For the linear simultaneous equations model the fact that $g(\delta)$ is differentiable is a consequence of the calculations reported in equations (3.7) and (3.8). Further, these calculations also imply that

$$(3.11) \qquad G = \text{plim}\ \frac{1}{T}\sum_{t=1}^{T} \partial g(X_t,\delta_o)/\partial\delta$$

Let $D_i = [\Pi_i, I_i]$, where $Y_i = Z\Pi_i + V_i$ is the reduced form equation for the right hand side endogenous variables in equation i and I_i is a selection matrix satisfying $ZI_i = Z_i$, and let $\tilde{D} = \text{diag}(D_1, \ldots, D_M)$. Then

$$(3.12) \quad \text{plim}(Z'X_i/T) = CD_i$$

Let $M_1 = -\text{plim}\frac{1}{T}\sum_{t=1}^{T}(\partial u_t'/\partial\delta) \otimes u_t'$ and $M_2 = -\text{plim}\frac{1}{T}\sum_{t=1}^{T} u_t' \otimes (\partial u_t'/\partial\delta)$.

Then from equtions (3.11) and (3.12) and $\partial u_t'/\partial\delta \times Z_t' = -\text{diag}(Z_t'X_{1t}, \ldots, Z_t'X_{Mt})$

$$(3.13) \quad G = -\begin{bmatrix} (I_M \otimes C)\tilde{D} \\ S(M_1 + M_2) \end{bmatrix}$$

The matrix $\tilde{D}$ is familiar from identification analysis of the simultaneous equations system under exclusion restrictions: see Hausman (1983, eq. (3.4)). The Lxq matrix $S(M_1 + M_2)$ is less familiar but can be interpreted. Suppose that G is partitioned into M matrices, where the ith matrix is the $(MK+L)xq_i$ matrix $\partial g/\partial\delta_i(\delta_o)$. The element of $S\cdot(u_t' \otimes u_t')$ corresponding to the restriction $\sigma_{ij} = 0$ is $u_{ti}u_{tj}$, so that the elements of this row of $S(M_1 + M_2)$ are zero except for the ith subvector which is $\text{plim}(u_j'X_i/T)$ and the jth subvector which is $\text{plim}(u_i'X_j/T)$. This row of $S(M_1 + M_2)$ involves the covariance of u_j with the right hand side variables of equation i and the covariance of u_i with the right hand side variables of equation j, which suggests that, as conjectured in Hausman and Taylor (1982), viewing residuals as instruments may yield identification criteria.

Formally, an assignment of residuals to equations is an L-tuple $a_m = (a_{m1},\ldots,a_{mL})$, where each element of a_m corresponds to a unique restriction $\sigma_{ij} = 0$. The element $a_{m\ell}$ is equal to i if residual j is assigned as an instrument to equation i and is equal to j if residual i is assigned as an instrument to equation j. For example, if the restriction corresponding to $\ell=1$ is $\sigma_{12} = 0$, then a_{m1} can either equal 1 or 2. If $a_{m1} = 1$ the second residual is assigned as an instrument to the first equation, and vice-versa if $a_{m1} = 2$. In general there are 2^L distinct possible assignments, which we will index by $m = 1,2,\ldots,2^L$.

For each assignment of disturbances as instruments, let U_{mi} be the matrix (possibly non-existent) of observations on disturbances which are assigned to equation i. Let $W_{mi} = [Z,U_{mi}]$ be the matrix of instrumental variables for equation i. Define

$$(3.14) \qquad N_{mi} = \text{plim}\ (W'_{mi}X_i/T)$$

and let $\tilde{N}_m = \text{diag}(N_{m1},\ldots,N_{mM})$. Now consider the matrix G. An obvious necessary condition for rank (G) = q is that G has to have at least q rows.

<u>Proposition 3.3</u>. If rank (G) = q then

$$\sum_{i=1}^{M} q_i \leq M\cdot K+L.$$

The predetermined variables Z provide K instruments for each equation and each covariance restriction provides one disturbance to be used as an instrument. Proposition 3.3 states that if the system is identified, the number of parameters in the system to be estimated can be no greater than the number of instruments available.

Now we suppose for the moment that G is square ($MK + L = q$) so that rank $(G) = q$ if and only if $\det(G)$ is non-zero. The following result give a formula for $\det(G)$ in terms of determinants which arise from assigning distrubances as instruments.

Lemma 3.1: For some 2^L-tuple of positive integers $(p_1,\ldots,p_{2^L})$

$$(3.15) \qquad \det(G) = \sum_{m=1}^{2^L} (-1)^{p_m}\det(\tilde{N}_m).$$

Proof: Let the rows of $S(M_1+M_2)$ be denoted by $s_\ell, \ell=1,\ldots,L$. Each ℓ corresponds to a restriction $\sigma_{ij}=0$ for some $i\neq j$. Further, each s_ℓ is a sum of two 1xq vectors, $s_{\ell i}+s_{\ell j}$ where $s_{\ell i}$ has $\text{plim}(u_j'X_i/T)$ for the subvector corresponding to δ_i and zeros for all other subvectors and $s_{\ell j}$ has $\text{plim}(u_i'X_j/T)$ for the subvector corresponding to δ_j and zeros for all other subvectors. We can identify $s_{\ell i}$ with an assignment of residual j to equation i and $s_{\ell j}$ with an assignment of residual i to equation j. We have

$$(3.16) \qquad -G = \begin{bmatrix} (I_M \otimes C)\tilde{D} \\ s_{1i} + s_{1j} \\ \vdots \\ s_{Li} + s_{Lj} \end{bmatrix},$$

where we drop an ℓ subscript on i and j for notational convenience. For each of the 2^L distinct assignments, indexed by m, let

$$(3.17) \qquad -\tilde{G}_m = \begin{bmatrix} (I_M \otimes C)\tilde{D} \\ \tilde{s}_m \end{bmatrix},$$

where $\tilde{s}_m$ is the Lxq matrix which has its ℓth row $s_{\ell i}$ if $a_{m\ell} = i$ or $s_{\ell j}$ if $a_{m\ell} = j$. The determinant of a matrix is a linear function of any particular row of the matrix: see Strang (1980, p. 156). It follows that if L = 1

$$(3.18) \qquad \det(-G) = \det(-\tilde{G}_1) + \det(-\tilde{G}_2).$$

Then induction on L gives

$$(3.19) \quad \det(-G) = \sum_{m=1}^{2^L} \det(-\tilde{G}_m)$$

Now consider $\tilde{G}_m$ for each m. The matrix $(I_M \otimes C)\tilde{D}$ is block diagonal, where the column partition corresponds to δ_i for i=1,...,M, and the ith diagonal block is plim $Z'X_i/T$. Further the ℓth row of $\tilde{S}_m$ consists of zeros except for the subvector corresponding to δ_i for $a_{m\ell}$= i, where plim$(u_j'X_i/T)$ appears. Then by interchanging pairs of rows of $\tilde{G}_m$, we can obtain $\tilde{N}_m$ from $\tilde{G}_m$. That is, $\tilde{N}_m = E_m\tilde{G}_m$, where E_m is a product of matrices which interchange a pair of rows of $\tilde{G}_m$. Note that E_m satisfies $E_m'E_m = I$, so that $\det(E_m) = (-1)^{p_m}$ for p_m equal to 1 or 2. It follows that $\det(\tilde{G}_m) = (-1)^{p_m}\det(\tilde{N}_m)$. Then since $\det(-G) = (-1)^q\det(G)$ and for each m $\det(-\tilde{G}_m) = (-1)^{p_m}\det(\tilde{N}_m)$

$$(3.20) \quad \det(G) = \sum_{m=1}^{2^L} \det(\tilde{G}_m) = \sum_{m=1}^{2^L} (-1)^{p_m}\det(\tilde{N}_m)$$

QED

Lemma 3.1 can be used to obtain strong necessary rank and order conditions for identification. We now drop the assumption that G is square.

<u>Proposition 3.4</u>: If rank(G) = q, then there must exist an assignment, m*, such that for each i, i=1,...,M, $\text{rank}(N_{m^*i}) = q_i$

Proof: If rank(G) = q, then there exists a q-dimensional square submatrix of G, denoted by $\bar{G}$, which is non-singular. The matrix $\bar{G}$ is obtained by deleting MK+L-q rows of G. Each row of $(I \otimes C)\tilde{D}$ which is deleted corresponds to ignoring an a variable in Z when considering instruments for an equation i. For each i let

$\bar{Z}_i$ denote the predetermined variables which remain as instruments for equation i after forming $\bar{G}$. Each row of $S(M_1 + M_2)$ deleted corresponds to ignoring a covariance restriction. Let $\bar{\ell} = 1,\ldots,\bar{L}$ index the remaining covariance

restrictions. For each assignment of distrubances as instruments, indexed as before by m, from the remaining covariance restrictions, let $\bar{W}_{mi} = (\bar{Z}_i, U_{mi})$ be the matrix of observations on the instrumental variables for equation i, and let

$\bar{N}_{mi} = \text{plim}(\bar{W}_{mi}'X_i/T)$ and $\tilde{\bar{N}}_m = \text{diag}(\bar{N}_{m1},\ldots,\bar{N}_{mM})$. Then from Lemma 3.1 it follows that

$$(3.21) \quad \det(\bar{G}) = \sum_{m=1}^{2^{\bar{L}}} (-1)^{p_m} \det(\tilde{\bar{N}}_m).$$

Then $\bar{G}$ non-singular implies $\det(\tilde{\bar{N}}_{\bar{m}}) \neq 0$ for some $\bar{m}$ and consequently rank $(\tilde{\overline{N}}_{\overline{m}}) = q$. Since $\tilde{\overline{N}}_{\overline{m}}$ is block diagonal,

$$(3.22) \quad \sum_{i=1}^{M} \text{rank}(\overline{N}_{\overline{m}i}) = \text{rank}(\tilde{\overline{N}}_{\overline{m}}) = q.$$

Each $\overline{N}_{\overline{m}i}$ has q_i columns so that $\text{rank}(\overline{N}_{\overline{m}i}) \leq q_i$ for i=1,...,M, and consequently equation (3.22) implies $\text{rank}(N_{\overline{m}i}) = q_i$ for i=1,...,M. Now let m* be an assignment of disturbances as instruments such that the covariance restrictions indexed by $\overline{\ell} = 1,\ldots,\overline{L}$ have disturbances assigned as the assignment indexed by $\overline{m}$ and for the other covariance restrictions the distrubances are assigned in any feasible fashion. Then for i=1,...,M

$$(3.23) \quad q_i \geq \text{rank}(N_{m^*i}) = \text{rank}(\text{plim}[Z:U_{m^*i}]'X_i/T)$$
$$\geq \text{rank}(\text{plim}[\overline{Z}_i:U_{\overline{m}i}]'X_i/T) = q_i$$

so that $q_i = \text{rank}(N_{m^*i})$. QED.

In order to check the rank of N_{mi}, it is useful to have an equivalent condition to rank $(N_{m^*i}) = q_i$ which is stated in terms of the structural parameters of the system. We assume without loss of generality, that i=1. Let $A = [B',\Gamma']'$ and let A_1 be the first column of A, which gives the coefficients of the first equation. Let the exclusion restrictions on the first equation be given by $\phi A_1 = 0$, where

ϕ is an $(M-1-q_1)$xMK selection matrix. For an assignment of disturbances as instruments, indexed by m, let Σ_{m1} be the rows of Σ corresponding to the disturbances assigned to the first equation.

Proposition 3.5: The rank of N_{m1} equals q_1 if and only if

$$\text{rank}\begin{bmatrix}\phi A_1 \\ \Sigma_{m1}\end{bmatrix} = M-1. \tag{3.24}$$

Proof: We drop the m subscript for notational convenience. Note that the first column of Σ_1 consists entirely of zeros, since to qualify as an instrument for the first equation a disturbance u_j must satisfy $E(u_1 u_j) = \sigma_{1j} = 0$. Let e_1 be an M dimensional unit vector with a one in the first position and zeros elsewhere. Then $\phi A_1 = 0$ implies $Fe_1 = 0$ where $F = (A'\phi', \Sigma_1')'$. Note that $\text{rank}(FB^{-1}) = \text{rank}(F)$. Also $FB^{-1}B_1 = Fe_1 = 0$ where B_1 is the first column of B, so that the first column of F is a linear combination of the other columns of F by $B_{11}=1$. Let Γ_1 be the rows of Γ corresponding to the excluded predetermined variables. Then $\phi AB^{-1} = [E_1', (B')^{-1}\Gamma_1']'$ where E_1 is an $(M-1-r_1)$xM matrix for which each row has a one in the position corresponding to a distinct excluded endogenous variable and zeros elsewhere. Let $(B^{-1})_1$ be the columns of B^{-1} corresponding to included right-hand side endogenous variables. Note that

$$(3.25)\qquad FB^{-1} = \begin{bmatrix} E_1 \\ \Gamma_1 B^{-1} \\ \Sigma_1 B\text{-}1 \end{bmatrix} .$$

Then row reduction of FB^{-1} using the rows of E_1, and the fact that the first column of FB^{-1} is a linear combination of the other columns imply

$$(3.26)\qquad \mathrm{rank}(FB^{-1}) = \mathrm{rank}\begin{bmatrix} \Gamma_1(B^{-1})_1 \\ \Sigma_1(B^{-1})_1 \end{bmatrix} + M-1-r_1 .$$

Now consider N_1. Note that for any $j \neq 1$,

$$(3.27)\qquad \mathrm{plim}\, u_j'X_1/T = [\mathrm{plim}(u_j'Y_1/T),\ \mathrm{plim}(u_j'Z_1/T)]$$

$$= [\mathrm{plim}(u_j'V_1/T),\ 0_1] = [\Sigma_j(B^{-1})_1, 0_1],$$

where 0_1 is a $1 \times s_1$ vector of zeros and Σ_j is the jth row of Σ. By C non-singular

$$(3.28)\qquad \mathrm{rank}(N_1) = \mathrm{rank}\left(\begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} D_1 \\ \mathrm{plim}\,(u_j'X_1/T) \end{bmatrix}\right)$$

$$= \mathrm{rank}\begin{bmatrix} \Pi_1 & I_1 \\ \Sigma_1(B^{-1})_1 & 0_1 \end{bmatrix} = \mathrm{rank}\begin{bmatrix} \Gamma(B^{-1})_1 & I_1 \\ \Sigma_1(B^{-1})_1 & 0_1 \end{bmatrix}.$$

By column reduction, using the columns of $[I_1'\ 0_1']'$, equation (3.28) implies

(3.29) $\text{rank}(N_1) = \text{rank}\begin{bmatrix}\Gamma_1(B^{-1})_1 \\ \Sigma_1(B^{-1})_1\end{bmatrix} + s_1.$

Then equations (3.26) and (3.29) imply

(3.30) $M\text{-}1\text{-rank}(F) = q_1 - \text{rank}(N_1),$

from which the conclusion of the proposition follows. QED.

For an assignment m, N_{mi} for i=1,...,M is the matrix of cross products of instrumental variables with right hand side variables. A necessary condition for identification is that there exists an assignment of disturbances as instruments to equations such that each cross product matrix of istruments with right hand side variables has full rank. Proposition 3.4 relates the rank of this matrix of covariances to the generalized rank condition of Fisher (1966, Theorem 4.62). A necessary condition for identification of a simultaneous equation system subject to covariance restrictions is that there exists an assignment of disturbances as instrumental variables to equations such that for each equation the generalized rank condition is satisfied for those covariance restrictions corresponding to disturbances assigned to the equation.

Proposition 3.4 implies the following order condition.

<u>Proposition 3.6</u>: If rank(G) = q, then there exists an instrument assignment such that for each i,i=1,...,M the number of disturbances assigned to equation i is at least q_i-K.

Proof: By proposition 3.4 there exists an instrument assignment such that for each i,i=1,...,M rank(N_{m*i}) = q_i. Then N_{m*i} must have at least q_i rows, and the conclusion follows from the fact that the number of rows of N_{m*i} equals K plus the number of disturbances assigned as instruments to equation i. QED

This order condition for system identification is stated in terms of the availability of enough instrumental variables to estimate each equation. Following Geraci's (1977) analysis of identification of simultaneous equations with errors in variables, we can give an algorithm for verifying whether there exists an assignment which yields enough instruments for each equation. As discussed by Geraci, this type of assignment problem is isomorphic to the mathematical problem of selecting a system of distinct representatives. We can apply Hall's theorem as follows: see Geraci (1977, p. 275).

For equation i,i=1,...,M, let $S_1^i,\ldots,S_{q_i-K}^i$ be q_i-K copies of the set of covariance restrictions $\sigma_{jk}= 0$ such that j=i or k=i. Each element of S_j^i corresponds to a potential instrument for equation i.

Let P be the ordered q-MK tuple $(S^1_1,\ldots,S^1_{q_1-K},\ldots,S^M_1,\ldots,S^M_{q_M-K})$. We can identify the selection of a covariance restriction $\sigma_{ik}=0$ from S^i_j as an assignment of disturbance k as an instrument for equation i. Hall's theorem then gives the following result.

Proposition 3.7: There exists an assignment of disturbances as instruments to equations such that for each i, i=1,...,M, at least q_i-K disturbances are assigned to equation i if and only if for each k, k=1,...,q-MK, the union of any k elements of P contains at least k distinct covariance restrictions.

Propositions 3.4 to 3.7 give stronger necessary conditions than any which have been presented in the econometrics literature for identification of a simultaneous equation system with covariance restrictions. Whether these conditions are strong enough to be sufficient for local identification remains an open question. Some insight into what is involved in the question of sufficiency can be gained by reconsidering Lemma 3.1. We suppose again that G is square. Let $\tilde{M}$ be the set of indices m which give assignments with the property that exactly q_i-K disturbances are assigned to the ith equation, for i=1,...,M. For each m in $\tilde{M}$, N_{mi} is square for i=1,...,M. Further, if m is not in $\tilde{M}$, then $\det(\tilde{N}_m) = 0$. This fact follows from $\text{rank}(\tilde{N}_m) = \sum_{i=1}^{M} \text{rank}(N_{mi})$, $\text{rank}(N_{mi}) \leq q_i$ for each i and $\text{rank}(N_{mi}) < q_i$ for some i, since if m is not in $\tilde{M}$, some equation must have less than q_i-K disturbances assigned to it which implies that N_{mi} has fewer than q_i rows. Lemma 3.1 then implies that

$$(3.31) \quad \det(G) = \sum_{m \text{ in } \tilde{M}} (-1)^{p_m} \left[\prod_{i=1}^{M} \det(N_{mi}) \right]$$

It follows immediately that if there is only one assignment possible, indexed as m*, which yields enough instruments for each equation, $\det(N_{m*i}) \neq 0$ for i=1,...,M is necessary and sufficient for local identification. The case of only one assignment giving enough instruments is essentially the case considered in Hausman and Taylor (1982) and Wegge (1965), where an ordering of equations is possible such that the identification of lower numbered equations does not depend on parameters of higher numbered equations.

In general, the index set $\tilde{M}$ contains more than one element, and $\det(N_{mi}) \neq 0$ for i=1,...,M and some m in $\tilde{M}$ will not guarantee that $\det(G) \neq 0$. However, it is our conjecture that if under the exclusion restrictions and the covariance restrictions there exists an m* such that almost everywhere $\det(N_{m*i}) \neq 0$ for i=1,...,M, then $\det(G) = 0$ will also be a measure zero event. This sufficiency result, if true, will be proved using much more of the structure of the matrix G than we have heretofore called upon.

We have not yet shown that if the disturbances are joint normally distributed then rank(G) = q is equivalent to non-singularity of the information matrix. The information matrix and G are also both important in estimation problems, so that it is natural to consider this question when discussing estimation. We now turn to estimation questions.

III. Estimation

In order to obtain asymptotic distribution results for the estimators which we consider for the linear simultaneous equations, it is necessary to strengthen our assumptions. We will assume that U_t has finite fourth moments, and that all sample averages of third and fourth order cross products of the elements of (Z_t, U_t) have a finite probability limit. Note that if U_t is normally distributed then U_t has finite moments of all orders, and in particular has finite fourth moments. For the moment we will postpone our discussion of the other implications of normality, and ignore the simplifications of the estimation problem which occurs if the disturbances are normally distributed.

Let $g_T(\delta) = (1/T) \sum_{t=1}^{T} g(X_t, \delta)$. Let $\tilde{\delta}_T$ be the GMM estimator obtained from solving

$$(3.32) \quad \min_{\delta \text{ in } \Theta} \; g_T(\delta)' \psi_T g_T(\delta),$$

where plim $\psi_T = \psi$ and ψ is positive definite. It is straightforward to show that $g_T(\delta)$ converges in probability to $g(\delta)$ uniformly in δ on any compact set, so that $g_T(\delta)'\psi_T g_T(\delta)$ converges in probability to $g(\delta)'\psi$ $g(\delta)$ uniformly in δ on any compact set. If we assume that Θ is compact, δ_o lies in Θ and δ_o is identified with respect to the moment condition functions $g(X_t,\delta)$, then plim $\tilde{\delta}_T = \delta_o$ follows from Amemiya (1973, Lemma 3). A more detailed examination of the conditions under which $\tilde{\delta}_T$ is consistent, which we do not undertake here, would probably reveal that compactness of Θ could be dispensed with. We will assume that plim $\tilde{\delta}_T = \delta_o$.

Let δ_o be in the interior of Θ. Then since plim $\tilde{\delta}_T = \delta_o$, with probability approaching one $\tilde{\delta}_T$ must satisfy

$$(3.33) \qquad 0 = \partial g_T(\tilde{\delta}_T)/\partial\delta' \psi_T g_T(\tilde{\delta}_T).$$

Multiply though by $\sqrt{T}$ and expanding the elements of $g_T(\tilde{\delta}_T)$ in a mean-value expansion around δ_o implies

$$(3.34) \qquad \partial g_T(\tilde{\delta}_T)/\partial\delta' \psi_T \partial g_T(\mathring{\delta}_T)/\partial\delta \sqrt{T}(\tilde{\delta}_T - \delta_o) = - \partial g_T(\tilde{\delta}_T)/\partial\delta' \psi_T \sqrt{T} g_T(\delta_o),$$

where $|\mathring{\delta}_T - \delta_o| < |\tilde{\delta}_T - \delta_o|$ and $\mathring{\delta}_T$ possibly differs from row to row of $\partial g_T/\partial\delta$. An appropriate central limit theorem implies that, since we have assumed U_t is i.i.d.,

$$(3.35) \qquad \text{plim} \sqrt{T} g_T(\delta_o) \overset{d}{\rightarrow} N(0,V),$$

where

$$(3.36)\quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{bmatrix},\ V_{11} = \Sigma \otimes C,$$

$$V_{12} = \text{plim}\ \frac{1}{T}\sum_{t=1}^{T} U_t'U_t \otimes Z_t'U_t\ S',\ V_{22} = SE(U_t'U_t \otimes U_t'U_t)S'.$$

It is straightforward to verify that $\delta_T \overset{P}{\to} \delta_o$ implies $\partial g_T(\delta_T)/\partial\delta \overset{p}{\to} G$. We will assume that δ_o is a regular point of $\partial g(\delta)/\partial\delta$ so that G has rank q. Then equation (3.34) implies $\sqrt{T}(\tilde{\delta}_T - \delta_o) = O_p(1)$, which together with $G'\Psi G$ non-singular implies

$$(3.37)\quad \sqrt{T}(\tilde{\delta}_T - \delta_o) = -(G'\psi G)^{-1}G'\psi\sqrt{T}g_T(\delta_o) + o_p(1).$$

For notational convenience, let $B=G'\psi$

Equation (3.37) implies that

$$(3.38)\quad \sqrt{T}(\tilde{\delta}_T - \delta_o) \overset{d}{\to} N(0,H)$$

where

$$(3.39)\quad H = (BG)^{-1}\ BVB'(G'B')^{-1}$$

is the asymptotic covariance matrix of $\tilde{\delta}_T$. We will discuss consistent estimation of H somewhat later.

Following Hansen (1982), if $\psi = V^{-1}$, or if MK+L = q, so that the asymptotic covariance matrix of $\tilde{\delta}_T$ is,

$$(3.40) \quad H^* = (G'V^{-1}G)^{-1},$$

then $\tilde{\delta}_T$ is an optimal GMM estimator, in the sense that H-H* is positive semi-definite. We will assume that MK+L>q, so that different choices of ϕ give different $\tilde{\delta}_T$. A general method of obtaining an optimal GMM estimator is to first solve the minimization problem (3.32) to obtain a consistent estimator of δ_o, which can then be used to form a consistent estimator V_T of V, and then to re-solve (3.32) with $\phi_T = V_T^{-1}$. Unfortunately, this general method of obtaining an optimal GMM estimator entails a second solution of the non-linear first order conditions for the problem (3.32). On computational grounds, it is useful to consider one step estimators of δ_o which have the optimal asymptotic covariance matrix H*. We could apply the general one step theorem for GMM estimators given in chapter 1, but for the problem of estimating the parameters of a simultaneous equations system with covariance restrictions, a more straightforward efficient instrumental variables estimator is often available. This estimator uses residuals as instruments in addition to the predetermined variables.

Since G is of full rank, proposition 3.4 implies that there exists an assignment of disturbances as instruments, with index m*, such that N_{m^*i} has rank q_i for i=1,...,M. Let $\tilde{u}_t$ be the 1xM residual vector with $\tilde{u}_{ti} = y_{it} - X_{it}\tilde{\delta}_T$, for i=1,...,M. Let $\bar{S}$ be an LxM^2 selection matrix which has the property that the ℓth element of $\bar{S}(u_t' \otimes \tilde{u}_t')$ corresponds to the ℓth covariance restriction $\sigma_{ij} = 0$, and if for the assignment m* disturbance j is assigned to equation i this element is $u_{ti}\tilde{u}_{tj}$, while if disturbance i is assigned to equation j, this element is $u_{tj}\tilde{u}_{ti}$. Let the MK+L vector of orthogonality condition functions $\bar{g}_T(X_t,\delta)$ be given by

$$(3.41) \qquad \bar{g}_T(X_t,\delta) = \begin{bmatrix} u_t' \otimes Z_t' \\ \bar{S}(u_t' \otimes \tilde{u}_t') \end{bmatrix}$$

and let

$$(3.42) \qquad \bar{g}_T(\delta) = \frac{1}{T} \sum_{t=1}^{T} \bar{g}_T(X_t,\delta).$$

These orthogonality condition functions have the usual form for instrumental variable estimators, since they consist of cross products of instrumental variables and residuals. In addition to the predetermined variables, the instrument list for each equation include residuals from $\tilde{u}_t$ assigned to according to the assignment m*. It will be notationally convenient for the discussion of asymptotic efficiency to stack the cross products of instruments and residuals as in equation (3.41) rather than stacking them according to equations.

The equivalence of $\bar{g}_T(\delta)$ to the usual instrumental variable orthogonality condition functions can be seen explicitly as follows.

Let $\tilde{U} = [\tilde{u}_1', \ldots, \tilde{u}_T']'$,

$\tilde{W} = [I_M \times Z,\ (I_M \otimes \tilde{U})\bar{S}']$, $y = (y_1', y_2', \ldots, y_M')'$ and

$\tilde{X} = \text{diag}([X_{11}', \ldots, X_{1T}']', \ldots, [X_{M1}', \ldots, X_{MT}']')$ and $\tilde{X}$ is MTxq. Let $u = y - \tilde{X}\delta$.

Then

$$(3.43) \qquad T\bar{g}_T(\delta) = \tilde{W}'u.$$

The matrix $\tilde{W}$ is the matrix of instruments for the stacked system $y = \tilde{X}\delta + u$. The selection matrix $\bar{S}'$ picks out the columns of $I_m \otimes \tilde{U}$ which correspond to residuals assigned as instruments for particular equations.

Instrumental variable estimators of δ_o can be obtained by solving the normal equations

$$(3.44) \qquad A_T \bar{g}_T(\hat{\delta}_T) = 0$$

to obtain, for $A_T'\tilde{W}'\tilde{X}$ non-singular,

$$(3.45) \qquad \delta_T = (A_T'\tilde{W}'\tilde{X})^{-1} A_T\tilde{W}'y,$$

where the matrix A_T is the (MK+L)xq linear combination matrix which forms the instruments $\tilde{W}A_T$ from the instrumental variables $\tilde{W}$. We will refer to estimators obtained as in equation (3.45) as augmented instrumental variable (AIV) estimators, where the word augmented refers to the fact that residuals are used as instruments, in addition to the predetermined variables.

We will assume that plim $A_T = A$. Also, our previous assumptions imply

$$(3.46) \quad \text{plim}\, \frac{\tilde{W}'\tilde{X}}{T} = -\,\text{plim}\frac{\partial \bar{g}_T}{\partial \delta} = \begin{bmatrix} (I_m \otimes C)\tilde{D} \\ -\bar{S}\ \text{plim}\ \frac{1}{T}\sum_{t=1}^{T} (\partial u_t'/\partial \delta \otimes \tilde{u}_t') \end{bmatrix}$$

$$= \begin{bmatrix} (I_m \otimes C)\tilde{D} \\ \bar{S}\ M_1 \end{bmatrix},$$

where the last equality follows from $\tilde{\delta}_T \xrightarrow{p} \delta_o$, which implies

$$\text{plim}(1/T) \sum_{t=1}^{T} -(\partial u_t'/\partial \delta \otimes \tilde{u}_t') = \text{plim}(1/T) \sum_{t=1}^{T} -(\partial u_t'/\partial \delta \otimes U_t') = M_1 .$$

Proposition 3.8: The sequence of (MK+L)xq matrices $\tilde{W}'\tilde{X}$ satisfies plim $\frac{\tilde{W}'\tilde{X}}{T}$ non-singular.

Proof: From equation (3.46), and the proof of proposition 3.4, it suffices to show that $\bar{S}M_1 = \tilde{s}_{m*}$, since then plim $\frac{\tilde{W}'\tilde{X}}{T} = -\tilde{G}_{m*}$. (The definitions of $\tilde{G}_{m*}$ and $\tilde{s}_{m*}$ are given during the discussion of identification). Consider the ℓth row of $\overline{S}M_1$, corresponding to the restriction $\sigma_{ij} = 0$, where, say, residual i is assigned to equation j. The ℓth row of $\overline{S}\ M_1$ is therefore given by

$\text{plim}\ \frac{\partial}{\partial\delta}\left[-\frac{1}{T}\sum_{t=1}^{T}\tilde{u}_{ti}u_{tj}\right] = -\ \text{plim}\ \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_{ti}(\partial u_{tj}/\partial\delta)'$, which by the definition of S_m is also the ℓth row of $\tilde{s}_{m*}$. QED

Let $\overline{G} = -\ \text{plim}\ \frac{\tilde{W}'\tilde{X}}{T}$. We assume that $A'\overline{G}$ is non-singular. Then $A_T'\tilde{W}'\tilde{X}$ is non-singular with probability approaching one as T gets large. Let $u_o = y - \tilde{X}\delta_o$. Then, following the usual IV analysis, δ_T satisfies

$$(3.47) \qquad \sqrt{T}(\hat{\delta}_T - \delta_o) = \left(\frac{A_T'\tilde{W}'\tilde{X}}{T}\right)^{-1} \frac{A_T'\tilde{W}'u_o}{\sqrt{T}} .$$

By the non-singularity of $A'\overline{G}$, $\text{plim}\left(\frac{A_T'\tilde{W}'\tilde{X}}{T}\right)^{-1} = (A'\overline{G})^{-1}$.

To obtain the asymptotic distribution of $\hat{\delta}_T$, we need to obtain the asymptotic distribution of $\frac{\tilde{W}'u_o}{\sqrt{T}} = \sqrt{T}\ \overline{g}_T(\delta_o)$. Here some care must be exercised, since $\tilde{W}$ depends on the estimated parameters $\tilde{\delta}_T$. In order to account for $\tilde{\delta}_T$, we will assume that $\tilde{\delta}_T$ satisfies equation (3.37), with $B' = G'\psi^{-1}$. Let $g_T(\delta)$ be partitioned as $g_T(\delta)' = (g_{T1}(\delta)', g_{T2}(\delta)')'$, where $g_{T1}(\delta_o) = (1/T) \sum_{t=1}^{T} U'_t \otimes Z'_t$ and

$g_{T2}(\delta_o) = (1/T) \sum_{t=1}^{T} S(U'_t \otimes U'_t)$. Let $B' = [B'_1, B'_2]$

also be partitioned conformably with $g_T(\delta)$. Note that $S\ (u_t' \otimes u_t') = \bar{S}(u'_t \otimes u'_t)$. Then we have

$$(3.48)\quad \frac{1}{\sqrt{T}} \bar{g}_{2T}(\delta_o) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{S}(U'_t \otimes \tilde{u}'_t)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} S(U'_t \otimes U'_t) + [\frac{1}{T} \sum_{t=1}^{T} \bar{S}\ (U'_t \otimes \partial u'_t/\partial\delta)]\ \sqrt{T}(\tilde{\delta}_T - \delta_o)$$

$$= \sqrt{T}\ g_{T2}(\delta_o) + \bar{S}M_2\sqrt{T}(\tilde{\delta}_T - \delta_o) + o_p(1),$$

(1).

where the last equality follows by the definition of M_2 and $\sqrt{T}\ (\tilde{\delta}_T - \delta_o) = O_p(1)$. Also, we have $\bar{g}_{T1}(\delta_o) = g_{T1}(\delta_o)$. Define the matrix

$$(3.49)\quad P = \begin{bmatrix} I_{MK} & 0 \\ \bar{S}M_2(B'G)^{-1}B'_1 & I_L + \bar{S}M_2(B'G)^{-1}B'_2 \end{bmatrix}.$$

Equation (3.48) and (3.37) imply

$$(3.50)\quad \frac{\tilde{W}'u_o}{\sqrt{T}} = \sqrt{T}\ \bar{g}_T(\delta_o) = P\sqrt{T}\ g_T(\delta_o) + o_p(1)$$

Then by $\sqrt{T}\ \bar{g}_T(\delta_o) = O_p(1)$ we have

(3.51) $\sqrt{T}\,(\hat{\delta}_T - \delta_o) = (A'\bar{G})^{-1}A'P\sqrt{T}\,g_T(\delta_o) + o_p(1)$,

and consequently the asymptotic covariance matrix of $\hat{\delta}_T$ is

(3.52) $\bar{H} = (A'\bar{G})^{-1}A'PVP'A(\bar{G}'A)^{-1}$.

Since $\bar{H}$ depends on A, it is desirable to consider the choice of A which minimizes $\bar{H}$, and gives the most efficient AIV estimator. Let $\bar{V}$ = PVP', so that $\bar{H} = (A'\bar{G})^{-1}A'\bar{V}A(\bar{G}'A)^{-1}$. Note that $\hat{\delta}_T$ can also be viewed as a GMM estimator with moment condition function $\bar{g}_T(X_t,\delta)$. The result of Hansen (1982) on an optimal GMM estimator cannot be directly applied, however, because due to the possibility of a singular P matrix $\bar{V}$ may not be invertible. Fortunately, the following generalization of Hansen's optimality result can be applied. Let $\bar{V}^-$ be any generalized inverse (g-inverse) of $\bar{V}$, and let $\bar{H}^* = (\bar{G}'\bar{V}^-\bar{G})^{-1}$ when this inverse exists

<u>Lemma 3.2</u>. If rank($\bar{G}$) = q and each column of $\bar{G}$ belongs to the column space of $\bar{V}$, then $\bar{G}'\bar{V}^-\bar{G}$ is non-singular and invariant with respect to choice of g-inverse. Further, $\bar{H} - \bar{H}^*$ positive semi-definite.

Proof: Since each column of $\bar{G}$ is an element of the column space of $\bar{V}$, there exists a matrix C such that $\bar{G} = \bar{V}C$. Let $\bar{V}^{1/2}$ be a symmetric square root of the positive semi-definite matrix $\bar{V}$. Then $\bar{G}'\bar{V}^-\bar{G} = C'\bar{V}\,\bar{V}^-\bar{V}C = C'\bar{V}C = (\bar{V}^{1/2}C)'(\bar{V}^{1/2}C)$. It follows that $\bar{G}'\bar{V}^-\bar{G}$ is invariant with respect to choice of g-inverse, and that $\text{rank}(\overline{G}'\overline{V}^{-}\overline{G}) =$ $\text{rank}(\overline{V}^{1/2}C) \geq$ $\text{rank}(\overline{V}^{1/2}\overline{V}^{1/2}C) = \text{rank}(\overline{V}C) = \text{rank}(\overline{G}) = q$. Also, rank $(A\overline{V}A') =$ $\text{rank}[(A\overline{V}^{1/2})\ \text{rank}(A\overline{V}^{1/2})'] = \text{rank}(A\overline{V})^{1/2} \geq \text{rank}(A\overline{V}^{1/2}\overline{V}^{1/2}C) =$ $\text{rank}(AG) = q$ by non singularity of AG so that $\overline{H}$ is positive definite. Then since $\overline{H}$ and $\overline{H}^*$ are positive definite, $\overline{H} - \overline{H}^*$ is positive semi-definite (p.s.d.) if and only if

$$(3.53)\qquad (\overline{H}^*)^{-1} - (\overline{H})^{-1} = \overline{G}'\overline{V}^{-}\overline{G} - \overline{G}'A'(A\overline{V}A')^{-1}A\overline{G}$$
$$= C'\overline{V}C - C'\overline{V}A'(A\overline{V}A')^{-1}A\overline{V}C$$
$$= C'\overline{V}^{1/2}[I - \overline{V}^{1/2}A'(A\overline{V}^{1/2}\overline{V}^{1/2}A')^{-1}A\overline{V}^{1/2}]\overline{V}^{1/2}C$$

is positive semi-definite which follows since the matrix in square brackets is indempotent. QED.

In order to apply Lemma 3.2 to AIV estimators, we need to show that $\overline{G}$ belongs to the column space of $\overline{V}$. The following lemma will imply this result, and is also important for determining the relationship between the optimal AIV estimator and the optimal GMM estimator obtained from solving (3.32) with $\phi = V^{-1}$.

<u>Lemma 3.3</u>: The matrices P, G and $\overline{G}$ satisfy $\overline{G} = PG$.

Proof: Let $G = [G_1', G_2']$ be partitioned conformably with P. Then

$$PG = \begin{bmatrix} I & 0 \\ \overline{S}M_2(B'G)^{-1}B_1' & I_L + \overline{S}M_2(B'G)^{-1}B_2' \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

$$= \begin{bmatrix} G_1 \\ G_2 + \overline{S}M_2(B'G)^{-1} - (\overline{S}M_2(B'G)^{-1})(B_1'G_1 + B_2G_2') \end{bmatrix} = \begin{bmatrix} -(I_M \otimes C)\tilde{D} \\ -\overline{S}M_1 \end{bmatrix}.$$

QED

We can now give the form of the optimal AIV estimator, and relate it to the optimal GMM estimator. Define the matrix $F = I_L + \overline{S}M_2(B'G)^{-1}B_2'$.

Proposition 3.9: Any AIV satisfying plim $A_T = \overline{V}^-\overline{G}$ for some symmetric g-inverse of $\overline{V}$ has asymptotic covariance matrix $A^* = (\overline{G}'\overline{V}^-\overline{G})^{-1}$ satisfying $\overline{H}-\overline{H}^*$ positive semi-definite. Further, for the covariance matrix $H^* = (G'V^{-1}G)^{-1}$ of an optimal GMM estimator, $\overline{H}^*-H^*$ is positive semi-definite, and if F is non-singular, then $\overline{H}^*=H^*$.

Proof: Note that $\overline{V} = PV^{1/2}(PV^{1/2})'$ where $V^{1/2}$ is a square root of V. Then by Rao (1973), the column space of $\overline{V}$ equals the column space of $PV^{1/2}$, which equals the column space of P by non-singularity of V. It follows next from Lemma 3.3 that there exists a matrix C with $\overline{G} = \overline{V}C$. Non-singularity of $\overline{G}'\overline{V}^-\overline{G}$ follows from Lemma 3.2. Setting $A = \overline{V}^-\overline{G}$ it follows that $A'\overline{G} = \overline{G}'\overline{V}^-\overline{G}$ is non-singular, and $\overline{H} = (\overline{G}'\overline{V}^-G)^{-1}\ \overline{G}'\overline{V}^-\overline{V}\overline{V}^-\overline{G}$ $(\overline{G}'\overline{V}^-\overline{G})^{-1} = \overline{H}^*C'\overline{V}^-\overline{V}\ \overline{V}\ \overline{V}^-GH^* = \overline{H}^*C'\overline{V}\ \overline{V}^-\overline{G}\ \overline{H}^* =$
$\overline{H}^*(\overline{H}^*)^{-1}\overline{H}^* = \overline{H}^*$. Positive semi-definiteness of $\overline{H}-\overline{H}^*$ follows Lemma 3.2.

Since $\overline{H}^*$ and H^* are positive definite, $\overline{H}^*-H^*$ will be positive semidefinite if $(H^*)^{-1} - (\overline{H}^*)^{-1}$ is. We have, using Lemma 3.3, $(H^*)^{-1} - (\overline{H}^*)^{-1} =$

$$G'V^{-1}G - G'P'(PV^{-1}P')^{-}PG = G'V^{-1/2}[I-V^{1/2}P'(PV^{1/2}V^{1/2}P')^{-}PV^{1/2}]V^{1/2}G$$

which is positive semi-definite by the fact that Rao (1973) implies the matrix in square brackets is idempotent. Also, if P is non-singular, then $\overline{H}^* = (G'P'(PVP)^{-1}PG)^{-1} = [G'P'(P')^{-1}V^{-1}P^{-1}PG]^{-1} = H^*$. The final conclusion of the proposition follows from the fact that P is block triangular with one diagonal block an identity matrix and the other diagonal block equal to F. QED.

We will refer to the optimal AIV estimator as augmented three stage least squares (A3SLS). The A3SLS estimator uses all of the instruments used by the 3SLS estimator (when it exists) but also uses as instruments the estimated residuals $\tilde{u}_t$ assigned as instruments according to the assignment m*. Whenever the 3SLS estimator exists, it is easily verified that the 3SLS estimator is an AIV estimator with

$$(3.54) \quad A' = [\tilde{D}'(\Sigma^{-1} \otimes I_K)\ ,\ 0],$$

where this choice of A is partitioned conformably with $g_T(\delta)$. It follows from proposition 3.9 that A3SLS will be more efficient than 3SLS. Further the condition that F is non-singular is likely to be satisfied in practice, so that A3SLS will be as efficient as the optimal GMM estimator. For instance, the case of most practical interest is a situation where the system of equations is identified by coefficient restrictions alone, so that we can obtain the initial estimator $\tilde{\delta}_T$ from two stage least squares (2SLS) or 3SLS estimation. In this case, the covariance restrictions are not used to form $\tilde{\delta}_T$ so that $B_2 = 0$. The non-singularity of F then follows immediately, since $B_2 = 0$ implies that $F = I_L$. Note that the A3SLS estimator has the same asymptotic covariance matrix, that of the optimal GMM estimator, whether 2SLS or 3SLS is used to form $\tilde{\delta}_T$. Intuitively, the A3SLS estimator adjusts for the initial estimator used to form the instrumental variables when forming the instruments $\tilde{W}A_T$ from $\tilde{W}$. Algebraically, this adjustment can be seen from the fact that, for F non-singular

(3.55) $\quad \text{plim}\ A_T = \bar{V}^{-}\bar{G} = (P')^{-1}V^{-1}G,$

The asymptotic distribution of the initial estimator affects the linear combination of instrumental variables through the matrix P.

To compute the A3SLS estimator, consistent estimators of P, $\bar{G}$ and V are required. Of course, $-\text{plim}\ \tilde{W}'\tilde{X}/T = \bar{G}$. Let

$$(3.56)\quad \tilde{V}_T = \frac{1}{T}\sum_{t=1}^{T} \bar{g}_T(X_t,\tilde{\delta}_T)\bar{g}_T(X_t,\tilde{\delta}_T)'$$

$$= \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \tilde{u}_t'\tilde{u}_t \otimes Z_t'Z_t & (\tilde{u}_t'\tilde{u}_t \otimes Z_t'\tilde{u}_t)\bar{S}' \\ \bar{S}(\tilde{u}_t'\tilde{u}_t \otimes \tilde{u}_t'Z_t) & \bar{S}(\tilde{u}_t'\tilde{u}_t \otimes \tilde{u}_t'\tilde{u}_t)\bar{S}' \end{bmatrix}.$$

Our assumptions on limits of sample averages of third and fourth order cross products of (Z_t,U_t), together with plim $\tilde{\delta}_T = \delta_0$, imply plim $\tilde{V}_T = V$. Since $\tilde{V}_T$ is an average outer product matrix $\tilde{V}_T$ is positive semi-definite. Let

$$\tilde{G}_T = (1/T)\sum_{t=1}^{T} \partial g(X_t,\tilde{\delta}_T)/\partial\delta \ ,\ \tilde{B}_T = \psi_T\tilde{G}_T \text{ and } \tilde{M}_{2T} = -(1/T)\sum_{t=1}^{T} \tilde{u}_t' \otimes \partial u_t'/\partial\delta .$$

Let

$\tilde{F}_T = I_L + \bar{S}\,\tilde{M}_{2T}(\tilde{B}_T'\,\tilde{G}_T)^{-1}\tilde{B}_{T2}$. Then plim $\tilde{B}_T = B$, plim $\tilde{G}_T = G$, plim $\tilde{M}_{2T} = M_2$ and plim $\tilde{F}_T = F$. Let

$$(3.57)\quad \tilde{P}_T = \begin{bmatrix} I_{MK} & 0 \\ \bar{S}\,\tilde{M}_{2T}(\tilde{B}_T'\,\tilde{G}_T)^{-1}\tilde{B}_{T1} & \tilde{F}_T \end{bmatrix},$$

so that plim $\tilde{P}_T = P$. We will assume throughout the rest of this chapter that P is non-singular. Finally, let $\tilde{\bar{V}}_T = \tilde{P}_T\tilde{V}_T\tilde{P}_T'$, so that plim $\tilde{\bar{V}}_T = \bar{V}_T$ which is non-singular. Then by $\bar{G}'\bar{V}^{-1}\bar{G}$ non-singular, with probability approaching one the A3SLS can be obtained from

$$(3.58) \qquad \delta_T = (\tilde{X}'\tilde{W}\,\tilde{\tilde{V}}_T^{-1}\tilde{W}'\tilde{X})^{-1}\,\tilde{X}'\tilde{W}\,\tilde{\tilde{V}}_T^{-1}\tilde{W}'y$$

The A3SLS estimator is an optimal instrumental variables estimator, where a consistent estimator of the asymptotic covariance matrix of $\tilde{W}'u_o/\sqrt{T}$ is given by $\tilde{\tilde{V}}_T$. A consistent estimator of the asymptotic covariance matrix of δ_T will be given by $\left(\frac{\tilde{X}'\tilde{W}}{T}\,\tilde{\tilde{V}}_T^{-1}\frac{\tilde{W}'\tilde{X}}{T}\right)^{-1}$.

The computations to obtain δ_T from (3.58) are only a little more laborious than those to obtain the 3SLS estimator, except for the computation of $\tilde{\tilde{V}}_T$. If the disturbances are normally distributed then $V_{12} = 0$, since all the third-order moments of a joint normal distribution vanish. Let $\tilde{\Sigma}_T = (1/T)\sum_{t=1}^{T} \tilde{u}_t'\tilde{u}_t$. Then under normality, a consistent estimator of V which is positive semi-definite is given by

$$(3.59) \qquad V_T = \begin{bmatrix} \tilde{\Sigma}_T \otimes (Z'Z/T) & 0 \\ 0 & \frac{1}{T}\sum_{t=1}^{T} \bar{S}(\tilde{u}_t'\tilde{u}_t \otimes \tilde{u}_t'\tilde{u}_t)\bar{S}' \end{bmatrix}$$

The situations where third order moments of U_T are zero are certainly more general than normally distributed disturbances, although if $V_{12} = 0$ is mistakenly assumed then the A3SLS estimator will be inefficient (although still consistent) and the asymptotic standard error estimators inconsistent.

The most important issues concerning normally distributed errors are those of identification and the asymptotic efficiency of the A3SLS estimator. Both issues can be addressed by examining the information matrix under covariance restrictions. In our discussion of the case of normally distributed errors we will draw heavily on the notation and results of Newey and Hausman (1982), where an expression for the information matrix in terms of the structural parameters is given.

Let $\tilde{B}_i = [(B^{-1})_i, O_i]'$ for i=1,...,M, where $(B^{-1})_i$ is a Mxr_i matrix of columns of B^{-1} corresponding to included endogenous variables in equation i and O_i is a Mxs_i matrix of zeros and let $\tilde{B}$ = $\text{diag}(\tilde{B}_1,...,\tilde{B}_M)$. Let $\tilde{\sigma} = \text{vec}\Sigma$ = $(\sigma_{11},...,\sigma_{M1},\sigma_{12},...,\sigma_{M2},...,\sigma_{MM})'$ and let σ^* = $(\sigma_{11},...,\sigma_{M1},\sigma_{22},...,\sigma_{M2},\sigma_{33},...,\sigma_{MM})'$ be a M(M+1)/2 vector of the unique elements of Σ. Note that σ^* is a parameterization of Σ which incorporates the symmetry restrictions, while $\tilde{\sigma}$ does not. Let R' be the M^2 x M(M+1)/2 defined in such a way that it maps σ^* into $\tilde{\sigma}$, i.e., $\tilde{\sigma} = R'\sigma^*$. Let S^* be a (M(M+1)/2-L)xM(M+1)/2 selection matrix such that $S^*\sigma^*$ consists of the non-zero elements of Σ. In Newey and Hausman (1982) it is shown that the information matrix for the simultaneous equations system with normally distributed disturbances and covariance restrictions imposed is given by

$$(3.60)\quad J = \begin{bmatrix} J_{\delta\delta'} & J_{\delta\sigma'} \\ J_{\sigma\delta} & J_{\sigma\sigma'} \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{B}(\Sigma^{-1} \otimes \Sigma)\tilde{B}' + \tilde{D}'(I_M \otimes C)\tilde{D}' & -\tilde{B}(\Sigma^{-1} \otimes I_M)R'S^{*\prime} \\ J_{\sigma\delta'} & \frac{1}{2} S^*R(\Sigma^{-1} \otimes \Sigma^{-1})R'S^{*\prime} \end{bmatrix}.$$

The non-singularity of $J_{\sigma\sigma'}$ follows from the fact that $R'S^{*\prime}$ is of full rank, since it is straightforward to show that the columns of $R'S^{*\prime}$ are orthogonal. Consider the matrix $J^* = J_{\delta\delta'} - J_{\delta\sigma'}J^{-1}_{\sigma\sigma'}J_{\sigma\delta'}$.

<u>Lemma 3.4</u> The information matrix J is non-singular if and only if J^* is non-singular.

Proof: This simple result follows directly from the calculations for the derivation of the partitioned inverse formula: see Theil (971). More directly, if J is non-singular then, since $J^* = [I,-J_{\delta\sigma'}J^{-1}_{\sigma\sigma'}]J[I,-J_{\delta\sigma'}J^{-1}_{\sigma\sigma'}]'$, J^* is non-singular. Also if J is singular, then there exists an $\chi \neq 0$ such that $J\chi = 0$. Let χ be partitioned conformably with $J, \chi' = (\chi_1', \chi_2')$. Then $J_{\sigma\delta'}\chi_1 + J_{\sigma\sigma'}\chi_2 = 0$ implies $\chi_2 = J^{-1}_{\sigma\sigma'}J_{\sigma\delta'}\chi_1$, which together with $J_{\delta\delta'}\chi_1 + J_{\delta\sigma'}\chi_2 = 0$ implies $J^*\chi_1 = 0$. Finally, if $\chi_1 = 0$ and $J\chi = 0$ then $\chi_2 = 0$, so that $\chi \neq 0$ implies $\chi_1 \neq 0$, and J^* is singular. QED

The non-singularity of J^* is thus equivalent to local identification of (δ_0, σ_0^*) (when this is a regular point). Also, when J^* is invertible,

J^{*-1} is the Cramer-Rao lower bound for the covariance matrix of an estimator of δ_o.

Next consider the matrix $G'V^{-1}G$. By the non-singularity of V, this matrix is non-singular if and only if rank(G) = q. Also, the inverse of this matrix, when it exists, is the asymptotic covariance matrix of the optimal GMM estimator. Therefore, if we can show that if the disturbances are normally distributed then $J^* = G'V^{-1}G$, it will follow that rank(G) = q is equivalent to non-singularity of the information matrix and that the optimal GMM estimator is asymptotically efficient. Intuitively, we expect that both of these propositions will hold under normality, since it is shown in Hausman and Taylor (1982) that the FIML estimator can be interpreted as an IV estimator, where residuals are used as instruments in addition to predetermined variables. Thus, FIML is a member of the class of GMM estimators we have considered, and the optimal GMM estimator should be asymptotically equivalent to FIML.

<u>Proposition 3.10</u>: If the disturbance vector U_t is normally distributed then $G'V^{-1}G = J^*$. Consequently, a regular point (δ_o, σ_o) is locally identified if and only if rank(G) = q and if (δ_o, σ_o) is identified then the optimal GMM estimator is asymptotically efficient.

Proof: For normally distributed disturbances,

$V^{-1}G = G_1'V_{11}^{-1}G_1 + G_2'V_{22}^{-1}G_2 = \tilde{D}'(\Sigma^{-1} \otimes C)\tilde{D} + G_2'V_{22}^{-1}G_2$. We will now draw on the results of Newey and Hausman (1982) (henceforth NH). Let Q be the $M^2 \times \frac{1}{2} M(M+1)$ matrix which has the same structure as R, except that in each row corresponding to σ_{ij}, $i > j$, the two non zero elements are 1/2 rather than 1 as in R. Let the vector U_t^* be the $\frac{1}{2} M(M+1) \times 1$ vector consisting of distinct products pairs of elements of U_t, with $U_t^* = (U_{t1}^2, U_{t2}U_{t1}, \ldots, U_{tM}U_{t1}, U_{t2}^2, \ldots, U_M^2)$. Note that $E(U_t^*) = \sigma^*$. Let S be a $L\times(1/2)M(M+1)$ selection matrix such that $\bar{S}(U_t' \otimes U_t') = SU_t^*$. Note that $S^*S' = 0$, since $S^*\sigma^*$ gives the unrestricted elements of σ^* and $S\sigma^*$ gives the restricted elements. It then follows from Richard (1975) and $E(SU_t^*) = 0$ that

(3.61) $$V_{22} = E(SU_t^*U_t^{*\prime}S') = Var(SU_t^*)$$

$$= S\ Var(U_t^*)S' = 2SQ(\Sigma \otimes \Sigma)Q'S'.$$

Let u_t^* be defined from u_t in the same manner that U_t^* was defined from U_t. Then it can easily be verified that

$$(3.62) \quad G_2 = S \text{ plim } \frac{1}{T} \sum_{t=1}^{T} \frac{\partial u_t^*}{\partial \delta}(\delta_o)$$

$$= S \text{ plim } \frac{1}{T} \sum_{t=1}^{T} [-2Q\text{diag}(U_t'X_{1t}, \ldots, U_t'X_{mt})]$$

$$= -2 \; SQ \text{ plim } \frac{(I_M \otimes U)'\tilde{X}}{T} = -2SQ(I_M \otimes \Sigma)\tilde{B}',$$

$$= -2 \; SQ(\Sigma \otimes \Sigma)Q'R(\Sigma^{-1} \otimes I_M)\tilde{B}'$$

where the last equality follows from use of the appendix of NH to show that

$$(3.63) \quad Q(\Sigma \otimes \Sigma)Q'R(\Sigma^{-1} \otimes I_M) = Q(I_M \otimes \Sigma).$$

As discussed in Richard (1975), $(R(\Sigma^{-1} \otimes \Sigma^{-1})R')^{-1} = Q(\Sigma \otimes \Sigma)Q'$. Let E be a singular matrix such that $R(\Sigma^{-1} \otimes \Sigma^{-1})R' = E^2$, so that $Q(\Sigma \otimes \Sigma) Q' = (E^{-1})^2$. Then from equations (3.61), (3.62) and equation (25) of NH we have

$$(3.64)\quad J^* - G'V^{-1}G = J^* - \tilde{D}'(\Sigma^{-1} \times C)\tilde{D} - G_2'V_{22}^{-1}G_2$$

$$= 2\tilde{B}(\Sigma^{-1} \times I_M)R$$

$$\times \left[(E^{-1})^2 - S^{*\prime}(S^*EES^{*\prime})^{-1}S^* - (E^{-1})^2S'(S(E^{-1})^2S')^{-1} S(E^{-1})^2\right]$$

$$\times R'(\Sigma^{-1} \times I_M)\tilde{B}'$$

$$= 2\tilde{B}(\Sigma^{-1} \times I_M)RE^{-1}$$

$$\times \left[I - ES^{*\prime}(S^*EES^{*\prime})^{-1}S^*E - E^{-1}S'(SE^{-1}E^{-1}S')^{-1}SE^{-1}\right]$$

$$\times E^{-1}R'(\Sigma^{-1} \times I_M)\tilde{B}'.$$

Let $[\cdot]$ denote the matrix in square brackets after the last equality sign in equation (3.64). Then since $S^*S' = 0$ and $S\,S^{*\prime} = 0$, $[\cdot]$ is idempotent. Further,

$$(3.65)\quad \text{rank}([\cdot]) = \text{trace}\,[\cdot]$$

$$= M(M+1)/2 - \text{trace}(ES'(S^*EES^{*\prime})^{-1}S^*E)$$

$$- \text{trace}(E^{-1}S'(SE^{-1}E^{-1}S')^{-1}SE^{-1})$$

$$= M(M+1)/2 - (M(M+1)/2 - L) - L = 0$$

and $[\cdot] = 0$ implies $J^* = G'V^{-1}G$. QED.

The estimation results we have presented can be summarized as follows. An initial consistent estimator of the parameters δ can be obtained by minimizing a quadratic form in the sample moment function $g_T(\delta)$, where $g_T(\delta)$ consists of cross products of predetermined variables and residuals and cross products of residuals with residuals corresponding to zero covariance restrictiaons. Using this initial consistent estimator, the A3SLS estimator can be formed, where A3SLS is an optimal IV estimator which uses residuals as instruments in appropriate equations, in addition to predetermined variables. If the matrix F is non-singular, as it is likely to be in practice, then the A3SLS estimator is as efficient as the optimal GMM estimator, and thus A3SLS (for non-singular F) will, asymptotically attain the Cramer-Rao lower bound.

IV. Testing Overidentifying Covariance Restrictions

When the system of equations is overidentified, the overidentifying restrictions can be tested. We will confine our discussion to tests based on the A3SLS estimator. Let the vector of A3SLS residuals be given by

(3.66) $\hat{u} = y - X\hat{\delta}_T$,

where $\hat{\delta}_T$ is the A3SLS estimator. Following Hansen (1982) and chapter 1, a general overidentifiability test statistic will be given by

(3.67) $m_T = \hat{u}'\tilde{W}\,\tilde{\bar{V}}_T^{-1}\tilde{W}'\hat{u}/T$,

which will be asymptotically distributed as chi-squared with MK + L -q degrees of freedom.

In the context of the classical simultaneous equations system, powerful tests of the covariance restrictions are particularly important. Covariance restrictions place restrictions on unobservables which may be somewhat less justified on the basis of economic theory than coefficient restrictions. The results of chapter 1 indicate that if the equation system is overidentified when the covariance restrictions are not imposed, (i.e., MK > q) then a more powerful test of the covariance restrictions than that based on m_T will be available.

In the terminology of Chapter 1, a test of the covariance restrictions is a test of whether the orthogonality conditions for the instrument residuals are satisfied. The form of such an optimal test can be obtained by applying Proposition 1.11 of Chapter 1. This test is based on the asymptotic distribution of

$$(3.68)\quad q_T = (1/\sqrt{T})[\bar{S}(I_M \otimes \tilde{U}')\hat{u} - \tilde{\bar{V}}_{T21}\tilde{\bar{V}}_{T11}^{-1}(I_M \otimes Z')\acute{u}]$$

where $\tilde{\bar{V}}_T$ is partitioned conformably with $\bar{g}_T(\delta)$, so that

$$(3.69)\quad \tilde{\bar{V}}_{T11} = (\tilde{P}_T\tilde{V}_T\tilde{P}_T')_{11} = \tilde{V}_{T11}$$

and

$$(3.70)\quad \tilde{\bar{V}}_{T21} = \bar{S}\,\tilde{M}_{2T}(\tilde{B}_T'\tilde{G}_T)^{-1}\tilde{B}_{T1}'\tilde{V}_{T11} + \tilde{F}_T\tilde{V}_{T21}$$

Note that even if the disturbances are normally distributed, implying plim $\tilde{V}_{T21} = V_{21} = 0$, the affect of estimated residuals causes the information contained in the cross-products of residuals and predetermined variables to be used in forming q_T.

From the discussion in the previous section and Chapter 1 it follows that a consistent estimator of Q, the asymptotic covariance matrix of q_T, is given by

$$(3.71)\quad \left(-\tilde{\tilde{V}}_{T21}\tilde{\tilde{V}}_{T11}^{-1}\ ,\ I_L\right)\left[\tilde{\tilde{V}}_T - \tilde{W}'\tilde{X}(\tilde{X}'\tilde{W}\tilde{V}_T^{-1}\tilde{X}'\tilde{W})^{-1}\tilde{X}'\tilde{W})\right]$$

$$\times\left(-\tilde{\tilde{V}}_{T21}\tilde{\tilde{V}}_{T11}^{-1}\ ,\ I_L\right)' = Q_T$$

From Proposition 1.1 of Chapter 1 it follows that if the system of equations is identified when the covariance restrictions are not imposed then

$$(3.72)\quad \bar{m}_T = q_T' Q_T^{-1} q_T$$

is asymptotically distributed as chi-squared with L degrees of freedom. If the system is overidentified without the covariance restrictions, then the test based on $\bar{m}_T$ will be locally more powerful, against an alternative where only the covariance restrictions are violated, than the test based on m_T.

A Hausman test based on the difference of the A3SLS and 3SLS estimators is probably easier to compute than $\bar{m}_T$. When this Hausman test has L degrees of freedom, it will be asymptotically equivalent to $\bar{m}_T$. Since it is likely that the number of coefficients in the simultaneous equations system is larger than L, this Hausman test should have L degrees of freedom, and its computation will require a generalized inverse. An appropriate method of forming this Hausman test will be considered in future research.

V. Conclusions

We have presented some strong necessary conditions for identification of a simultaneous equations system with restrictions on the disturbance covariances. These conditions involve an equation by equation examination of rank conditions, once the order condition, involving an assignment of residuals as instruments to equations, is satisfied. Furthermore, the rank condition is of the simple form given in Fisher (1966) as the generalized rank condition. The question of whether our necessary conditions are sufficient for local identification is unresolved and remains a topic for future research.

To estimate the simultaneous equations parameters we have obtained an augmented three-stage least squares estimator which is more efficient than 3SLS. The full characterization of those situations with normal disturbances in which A3SLS is fully efficient is an open question. We conjecture that it will only be a set of measure zero in the parameter space on which A3SLS could fail to be fully efficient. The amount of precision gained by using A3SLS rather than 3SLS also remains an open question, although Rothenberg (1973) has investigated the extent of efficiency gain obtained in some situations.

Finally, we have given an optimal method of testing covariance restrictions based on the A3SLS estimator. This test of covariance restrictions has potential applications beyond the classical simultaneous equations system. Properly modified, tests of covariance restrictions will be relevant for some rational expectations models, such as that of Mishkin (1982).

BIBLIOGRAPHY

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

Amemiya, T.: "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, 41 (1973), 997-106.

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

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

Eichenbaum, M. S., L. P. Hansen and K. J. Singleton: "Testing Restrictions in Nonlinear Rational Expectations Models." Manuscript, 1982.

Engle, R. F., D. F. Hendry and J. F. Richard: "Exogeneity," Econometrica, 51 (1983), 277-304.

Fisher, F.: The Identification Problem in Econometrics. New York: McGraw-Hill, 1966.

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

Geraci, V. J.: "Identification of Simultaneous Equation Models with Measurement Error," Journal of Econometrics, 3 (1975), 57-60.

Hall, P.: "On Representatives of Subsets," Journal of the London Mathematical Society, 10 (1935), 26-30.

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

Hausman, J. A.: "Specification Tests in Econometrics," Econometrica, 46 (1978), 1251-1272.

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

Hausman, J. A. and W. K. Newey: "The Informaion Matrix for the Linear Simultaneous Equations System." Manuscript, MIT, 1981.

Hausman, J. A. and W. E. Taylor: "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation," Bell Laboratories Economic Discussion Paper, 1982.

Hausman, J. A.: "Specification and Estimation of Simultaneous Equations Modls," chapter 7 in Griliches, Z. and M. D. Intriligator (eds), Handbook of Econometrics, Vol I., Amsterdam: North-Holland Publishing Co., 1983.

Hoadley, Bruce: "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," Annals of Mathematical Statistics, 42 (1971), 1977-1991.

Holly, A.: "A Remark on Hausman's Specification Test," Econometrica, 50 (1982), 749-759.

Hurd, M.: "Estimation in Truncated Samples When There is Heteroscedasticity," Journal of Econometrics, 11 (1979), 247-258.

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

Koopmans, T. C., H. Rubin and R. B. Leipnik: "Measuring the Equations Systems of Dynamic Economics," in Koopmans, T. C. (ed), Statistical Inference in Dynamic Economic Models (Cowles Commission Monograph 10) New York: John Wiley and Sons, 1950.

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

Lee, Lung-fei: "A Specification Test for the Normality Assumption for the Truncated and Censored Tobit Models." Manuscript, Univ. of Minnesota, 1981.

Malinvaud, E.: Statistical Methods of Econometrics (2nd ed.). Amsterdam: North-Holland Publishing Co., 1970.

Manski, C. F. and D. McFadden: "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in Manski, C. F. and D. McFadden (eds), Structural Analysis of Discrete Data with Applications, Cambridge (Massachusetts): MIT Press, 1981.

McFadden, D.: "Asymptotic Properties of Maximum Likelihood and Nonlinear Least Squares Estimators." Manuscript (Lecture Notes), 1980.

Mishkin, F. S.,: "Does Anticipated Monetary Policy Matter? An Econometric Investigation," Journal of Political Economy 90 (1982), 22-51.

Nelson, F. D.: "A Test for Misspecification in the Censored Normal Model," Econometrica, 49 (1981), 1317-1329.

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

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

Richard, J. F.: "A Note on the Information Matrix of the Multivariate Normal Distribution," Journal of Econometrics, 3 (1975), 57-60.

Rothenberg, T. J.: "Identification in Parametric Models," Econometrica, 39 (1971), 577-592.

Rothenberg, T. J.: Efficient Estimation with a priori Information, (Cowles Foundation Monograph 23). New Haven: Yale University Press, 1973.

Ruud, P.: Misspecification in Limited Dependent Variable Models, Unpublished Ph.D. dissertation, MIT, 1981.

Strang, G.: Linear Algebra and its Applications (2nd ed.). New York: Academic Press, 1980.

Wegge, L.: "Identifiability Criteria for a System of Equations as a Whole," The Australian Journal of Statistics, 7 (1965), 67-77.

White, H.: "Nonlinear Regression on Cross-Section Data," Econometrica, 729-746.

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

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