2-1-1974

The Moments of the 3SLS Estimates of the Structural Coefficients of a Simultaneous Equation Model

J. D. Sargan

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 370

THE MOMENTS OF THE 3SLS ESTIMATES OF THE STRUCTURAL COEFFICIENTS OF A SIMULTANEOUS EQUATION MODEL

J. D. Sargan

February 25, 1974

THE MOMENTS OF THE 3SLS ESTIMATES OF THE STRUCTURAL COEFFICIENTS OF A SIMULTANEOUS EQUATION MODEL*

by

J. D. Sargan

1. Introduction

This article is concerned with the maximal finite moment exponent for the 3SLS estimates of a set of structural coefficients. This is defined for any estimator $\hat{\beta}$ as the supremum of r such that $E(|\beta|^r) < \infty$. The strategy in this paper is to establish an upper bound on the maximal exponent through Theorem 1; then to establish a lower bound through Theorem 2. The paper considers the general case where there are general linear restrictions which may connect the coefficients of different equations in the model. For this general case the results of the paper may not be enough to exactly specify the maximal exponent. However in the important special case where the restrictions are separable, in the sense that each constraint involves only the coefficients of a single equation of the model, it is established that the maximal exponent for the coefficients of any equation is the same as that for the corresponding 2SLS estimates. That is, the maximal exponent is equal to the number of overidentifying restrictions, or to the difference between the number of restrictions on the equation minus the number of endogenous variables in the model. All these results are derived in the case where the 2SLS estimates of the structural equation error variance matrix have been modified so that the ratio of the

*The research described in this paper was undertaken from grants from the National Science Foundation and from the Ford Foundation.

largest to the smallest latent root is bounded. The last section considers the case where this change is not made to the conventional definition of the estimators.

2. The Model and an Upper Bound Theorem

The general linear model will be written in the form

$$AX' = BY' + CZ' = U' ,$$

where Y and U are nxT matrices of stochastic variables, and Z is an mxT matrix of non-stochastic exogenous variables. B is a square non-singular matrix of coefficients, $A = (B : C)$ and $X = (Y : Z)$, and $E(U) = 0$. Writing u_t' for a row of U , u_t is independent of u_s , $s \neq t$, with an absolutely continuous joint distribution for the element of u_t independent of t , with finite variance matrix Ω_u . Writing $A' = (a_1, a_2, a_3, \ldots, a_n)$, we write $(\text{vec } A') = (a_1', a_2', a_3', \ldots, a_n')$, as an $n(m+n)$ vector. We then assume a general set of restrictions of the form

(1) $$\Phi \text{ vec } A = \phi ,$$

where Φ is $rxn(n+m)$ of rank r .

The 3SLS estimates can be regarded as minimizing

$$\text{tr}(\Psi ARA') = (\text{vec } A)'(\Psi \otimes R)\text{vec } A$$

subject to the constraints (1), where $R = (X'Z)(Z'Z)^{-1}(Z'X)/T$. Ψ would normally be taken to be the inverse of $\hat{\Omega}_u$, where $\hat{\Omega}_u$ is the 2SLS estimate of Ω_u . However this is a difficult choice of Ψ , and it is much easier to deal with the case where Ψ is non-stochastic, or has the ratio of its largest and smallest latent roots bounded. Such a matrix might be derived from $\hat{\Omega}_u^{-1}$ by reducing any latent root which was greater than a fixed rather large multiple μ of the smallest latent root. If the ratio of largest to smallest latent root of the population Ω_u is in fact smaller than this μ then the possibility that the bound would have to be applied to any sample $\hat{\Omega}_u$ would tend to zero with T, and the resulting estimates would be asymptotically equivalent to the usual estimates using $\Psi = \hat{\Omega}_u^{-1}$.

Given Ψ we can state the general estimator by parameterizing the constraints. Since Φ is of full rank we can find a permutation matrix Π such that $\Phi\Pi = (\Phi_1 : \Phi_2)$, and Φ_1 is square non-singular. Writing $(\Pi' \text{ vec } A)' = (\alpha_0' : \alpha')$ so that $\Phi_1\alpha_0 + \Phi_2\alpha = \phi_0$. Then
$\alpha_0 = \Phi_1^{-1}\phi_0 - \Phi_1^{-1}\Phi_2\alpha$.

Thus

$$\text{vec } A = \Pi\begin{pmatrix}\alpha_0\\ \alpha\end{pmatrix} = \Pi\begin{pmatrix}\Phi_1^{-1}\phi_0\\ 0\end{pmatrix} + \Pi\begin{pmatrix}-\Phi_1^{-1}\Phi_2\\ I\end{pmatrix}\alpha ,$$

and we write this as $\text{vec } A = k_0 + K\alpha$. This is only one out of many ways of parameterizing the constraints, but it is one that is easily programmed for computer and it is possible to speed up computing by using the specialized form of the K matrix and k_0 vector.

We now minimize

$$k_0'(\Psi \otimes R)k_0 + 2k_0'(\Psi \otimes R)K\alpha + \alpha'(K'(\Psi \otimes R)K)\alpha$$

unconstrainedly with respect to α, where α is an N vector, $N = n(n+m) - r$. Thus $\hat{\alpha} = -(K'(\Psi \otimes R)K)^{-1}K'(\Psi \otimes R)k_0$, and $\text{vec}\,\hat{A} = k_0 + K\hat{\alpha}$. The model is identified if $\underset{T\to\infty}{\text{p lim}}(K'(\Psi \otimes R)K)$ is positive definite and the asymptotic error variance matrix can be estimated as $K(K'(\Psi \otimes R)K)^{-1}(K'(\Psi\hat{\Omega}_u\Psi \otimes R)K)(K'(\Psi \otimes R)K)^{-1}K'$. This takes the conventional form $K(K'(\Psi \otimes R)K)^{-1}K'$, if $\underset{T\to\infty}{\text{p lim}}\,\Psi = \Omega_u^{-1}$.

Note that the parameterization method gives exactly the same estimates as would be obtained by a Lagrange multiplier approach, but saves on both computer time and storage space.

In order to set an upper bound to the maximal finite moment exponent we consider an arbitrary constant linear function of the elements of $(\hat{\alpha} - \alpha)$ of the form

$$h'(\hat{\alpha} - \alpha) = h'(K'(\Psi \otimes R)K)^{-1}K'(\Psi \otimes (X'Z)(Z'Z)^{-1})\text{vec}(U'Z)/T .$$

We can express this as a function of $\hat{P} = (Y'Z)(Z'Z)^{-1/2}/\sqrt{T}$, and $W = \Omega_v^{-1/2}(Y'Y - (Y'Z)(Z'Z)^{-1}(Z'Y))\Omega_v^{-1/2}/T$, in so far as Ψ depends on this and $\hat{P}$. Writing $p = \text{vec}\,\hat{P}$, and defining w as a vector whose elements are the upper triangle of the symmetric matrix W, we can think of $h'(\hat{\alpha} - \alpha)$ as a function of p and w.

Considering the integral defining the k^{th} absolute moment of $h'(\hat{\alpha} - \alpha)$, the singular points of the integrand are points where $\det(K'(\Psi \otimes R)K) = 0$. We write $\psi(p,w) = \det(K'(\Psi \otimes R)K)$, and refer to the set of points in (p,w) space where $\psi(p,w) = 0$ as the zero set. The next theorem depends upon the idea that if the integral defining the moment diverges in the neighborhood of some point of the zero set, then the moment will not exist.

In general the zero-set is made up of one or more differential manifolds [2]. For each such manifold we can define one or more patches on which we can parametrize the points of the manifold in the form $p = p^*(\theta)$, $w = w^*(\theta)$, where $\psi(p^*(\theta), w^*(\theta)) = 0$ for all points θ lying in some open set in θ space. In our particular case it turns out that the differential manifolds are cylinders in the sense that effectively only the p vector is constrained by $\psi(p,w) = 0$, so that the differential manifold is parameterized in the form $p = p^*(\theta)$, w arbitrary. For the next theorem it is only necessary to consider such a parameterization in some neighborhood of a point $(p_0, w_0) = (p^*(\theta_0), w_0)$ in (p,w) space, where (p_0, w_0) belongs to the zero set.

Theorem 1. If for some point (p_0, w_0) in the zero set there exists a closed neighborhood of radius δ, $n(p_0, w_0, \delta)$ such that:

(A) $\Psi(p,w)$ is continuous on the neighborhood and $\Psi(p_0, w_0)$ is positive definite,

(B) $K'(\Psi \otimes R)K$ is of rank $N-1$ at (p_0, w_0), and δ is sufficiently small that this is the minimal rank attained at any point of the neighborhood. If η_0 is the unique vector such that $K'(\Psi \otimes R)K\eta_0 = 0$ at (p_0, w_0) then $h'\eta_0 \neq 0$,

(C) There exists a differential manifold contained in the zero-set, which can be parameterized in the form $p = p^*(\theta_1)$, $w = \theta_2$, such that $\psi(p^*(\theta_1), \theta_2) = 0$ for all points (θ_1, θ_2) lying in some open set in θ space including the point (θ_0, w_0)

and for all (θ_1, θ_2) such that $(p^*(\theta_1), \theta_2)$ lies in $n(p_0, w_0, \delta)$, where $p_0 = p^*(\theta_0)$. For points in the neighborhood $p^*(\theta_1)$ has continuous second derivatives and the first derivative matrix is of rank q, where q is the dimension of θ_1,

(D) The probability density of the stochastic variables exists and is continuous at (p_0, w_0),

(E) The vector z_0 defined below is non-zero,

then $E(|h'(\hat{\alpha} - \alpha)|^k)$ is unbounded if $k \geq mn - q$.

<u>Proof</u>. A sufficient condition that $\det(K'(\Psi \otimes R)K) = 0$ is that for some non-zero vector θ^*

$$(I \otimes (Z'Z)^{-1/2}(Z'X))\theta^* = 0 . \tag{2}$$

Since (2) does not involve w, but only p, if we can parameterize part of the zero set from (2) in the form $p = p^*(\theta_1)$, then such points are in the zero set for all w. The points of the zero set which do not belong to such differential cylinders will be those where $\det \Psi = 0$. From assumption (A) we can choose δ so small that such points do not occur in the neighborhood, and so all points of the zero set can be assumed to satisfy (2) for some θ^*. Normalizing θ^* so that its largest element is one, we will later consider the possibility of solving (2) for $\hat{P}$ as a function of the remaining elements of θ^*, which will then make up the the vector θ_1 referred to in assumption (C).

From assumption (B) N-1 of the latent roots of $K'(\Psi \otimes R)K$ are non-zero throughout $n(p_0, w_0, \delta)$ and from continuity and compactness we can bound them below by a positive constant. Write $K'(\Psi \otimes R)K = LDL'$ where D is a diagonal matrix of non-negative latent roots and L is the corresponding matrix of latent vectors. Now write d^2 for the smallest latent root and π^* for the corresponding latent vector, and write D^* for D^{-1} with its largest element d^{-2} replaced by zero. Also define $z = T^{-1/2}(\Psi^{1/2} \otimes (Z'Z)^{-1/2}(Z'X))K\pi^*$, so that

$$
\begin{aligned}
Th'(\hat{\alpha} - \alpha) &= h'LD^{-1}L'(K'[\Psi \otimes (X'Z)(Z'Z)^{-1}])\text{vec}(U'Z) \\
&= h'LD^*L'K'(\Psi \otimes (X'Z)(Z'Z)^{-1})\text{vec}(U'Z) \qquad (3)\\
&\quad + (h'\pi^*)z'[\Psi^{1/2} \otimes (Z'Z/T)^{-1/2}]\text{vec}(U'Z)/d^2 .
\end{aligned}
$$

We have defined z so that

$$z'z = \pi^{*\prime}(\Psi \otimes R)\pi^{*\prime} = d^2 . \qquad (4)$$

Now writing the local parameterization for p as $p = p^*(\theta_1)$, we define $\theta^*(p)$ as the value of θ_1 minimizing $\|p - p^*(\theta_1)\|^2$. It is not difficult to show that for $\|p - p_0\| \leq \delta$ and δ sufficiently small $\theta^*(p)$ is uniquely determined as a function of p with continuous first derivatives using assumption (C). Define $p^+(p) = p^*(\theta^*(p))$, then $p^+(p)$ is well defined with continuous first derivatives in the neighborhood. The first order conditions for a minimum are

$$(p - p^+(p))' \frac{\partial p^*}{\partial \theta_1} = 0 \qquad (5)$$

where the derivatives are taken at $\theta^*(p)$. From (C) $\partial p^*/\partial\theta_1$ is of rank

q , and so we can define the idempotent matrix

$$Q(\theta_1) = I - \left(\frac{\partial p^*}{\partial \theta_1}\right)\left(\frac{\partial p^{*\prime}}{\partial \theta_1}\frac{\partial p^*}{\partial \theta_1}\right)^{-1}\left(\frac{\partial p^*}{\partial \theta_1}\right)' .$$

By suitable choice of linear restrictions on the elements of $H(\theta_1)$ we can uniquely define it from the factorization

$$H(\theta_1)H(\theta_1)' = Q(\theta_1) \quad \text{and} \quad H(\theta_1)'H(\theta_1) = I . \tag{6}$$

Now define $z^*(p) = H(\theta^*(p))'(p - p^+(p))$, and note that we have that $Q(\theta^*(p))(p - p^+(p)) = p - p^+(p)$ from the definition of $Q(\theta_1)$ and equation (5). Thus $p - p^+(p) = H(\theta^*(p))z^*(p)$, using (6).

Consider now the change of variables from p to (θ_1, z^*) defined by the equation

$$p = p^*(\theta_1) + H(\theta_1)z^* . \tag{7}$$

It is clear that $\theta^*(p) = \theta_1$ satisfies the equation (7), and from the uniqueness of the definition of $\theta^*(p)$ given that θ is sufficiently small, it follows that (7) defines a one to one mapping with continuous first derivatives. Then

$$\frac{\partial p}{\partial \theta_1} = \frac{\partial p^*}{\partial \theta_1} + \frac{\partial H}{\partial \theta_1} z^*$$

$$\frac{\partial p}{\partial z^*} = H .$$

The Jacobian of the transformation tends to

$$\left| \det \begin{pmatrix} \partial p^*/\partial\theta_1 \\ H \end{pmatrix} \right|$$

as $z^* \to 0$, and $\partial p^*/\partial\theta_1$ and H are mutually orthogonal and each of full rank at $p = p_0$. So for sufficiently small λ the Jacobian is bounded below in the neighborhood. Consider now the integral of the k^{th} moment on a set ω defined by $\|\theta_1 - \theta_0\| \leq \epsilon_1$, $\|w - w_0\| \leq \epsilon_2$, $\epsilon_4^2 \leq z'^*(p)z^*(p) \leq \epsilon_3^2$. ϵ_1 , ϵ_2 , and ϵ_3 are chosen so that the set lies within the neighborhood $n(p_0, w_0, \lambda)$. Let f_m be the minimum of the possibility density and J_m be the minimum of the Jacobian on ω . Then the integral of the moment satisfies

$$I \geq J_m f_m \int_\omega |h'(\hat{\alpha} - \alpha)|^k d\theta_1 dz^* dw \ .$$

Now

$$(8) \quad z(p,w) = z(p^*(\theta_1) + H(\theta_1)z^*, w) = z(p^*(\theta_1), w) + (\partial z/\partial p)H(\theta_1)z^* \ .$$

$\partial z/\partial p$ is computed at a point in p space between $p^*(\theta_1)$ and $p^*(\theta_1) + H(\theta_1)z^*$. Note that from (4) $z(p^*(\theta_1), w) = 0$ since d^2 is then the smallest latent root of a singular matrix and is so equal to zero. So $z(p,w) = (\partial z/\partial p)H(\theta_1)z^*$.

Also $\mathrm{vec}(U'Z) = \mathrm{vec}(A(X'Z)) = T\ \mathrm{vec}(B\hat{P}M^{1/2} + CM)$ where $M = Z'Z/T$. If Ψ_0 is the value of Ψ at (p_0, w_0) , Z_p is the value of $\partial z/\partial p$ at the same point and $d_0 = \mathrm{vec}(BP_0 + CM^{1/2})$, where $\mathrm{vec}\ P_0 = p_0$, we can define

$$v(p,w) = d^2 h'(\hat{a} - a)/(z^{*\prime}z^*)^{1/2} - (h'\eta_0)z^{*\prime}H'(\theta_0)z_p'(\Psi_0^{1/2} \otimes I)d_0/(z^{*\prime}z^*)^{1/2}$$

$$= d^2/(z^{*\prime}z^*)^{1/2}[h'(LD^*L')K'(\Psi \otimes (X'Z)(Z'Z)^{-1}/T)\text{vec}(U'Z)]$$

$$+ z^{*\prime}/(z^{*\prime}z^*)^{1/2}[h_1'\eta^*(p))H'(\theta_1)(\partial z/\partial p)'(\Psi^{1/2} \otimes I)(\text{vec}(B\hat{P} + CM^{1/2})$$

$$- (h'\eta_0)H'(\theta_0)Z_p'(\Psi_0^{1/2} \otimes I)d_0] .$$

Now $d^2 = z'z = z^{*\prime}(H'(\theta_1)(\partial z/\partial p)'(\partial z/\partial p)H(\theta_1))z^* \leq \lambda_M(z^{*\prime}z^*)$ where λ_M is an upper bound on the largest latent root of $H'(\theta)(\partial z/\partial p)'(\partial z/\partial p)H(\theta)$ in the neighborhood. Since the vector $z^*/(z^{*\prime}z^*)^{1/2}$ is bounded it follows from continuity that $|v(p,w)|$ is bounded above by a quantity v_0 which tends to zero with δ . Define

$$z_0 = (h'\eta_0)H'(\theta_0)Z_p'(\Psi_0^{1/2} \otimes I)d_0 , \quad \text{then}$$

$$\frac{d^2|h'(\hat{a} - a)|}{(z^{*\prime}z^*)^{1/2}} \geq \frac{|z^{*\prime}z_0|}{(z^{*\prime}z^*)^{1/2}} - v_0 , \quad \text{if} \quad |(z^{*\prime}z_0)| > v_0(z^{*\prime}z^*)^{1/2} ,$$

$$\geq 0 \qquad\qquad , \quad \text{otherwise.}$$

By taking δ sufficiently small we ensure that $v_0^2 < (z_0'z_0)$. Thus

$$I \geq J_m f_m k_1 \epsilon_1^a \epsilon_2^{\frac{1}{2}n(n+1)} / \lambda_M^k \int (z^{*\prime}z_0/(z^{*\prime}z^*)^{1/2} - v_0)^k (z^{*\prime}z^*)^{-\frac{1}{2}k} dz^*$$

where the range of integration is restricted to points where $|z^*z_0| \geq v_0(z^{*\prime}z^*)^{1/2}$, and k_1 is the constant of integration obtained because we have integrated out both θ_1 and w over appropriate hyper-spheres. Now first rotate the z^* axes by introducing $z^+ = L^*z^*$, where

L^* is an orthogonal matrix whose first row is $z_0'/(z_0'z_0)^{1/2}$, and then transform the z^+ variables by the usual polar coordinates transformation of the form $R^* = (z^{+'}z^+)^{1/2}$, $z_i^+ = R^* \prod_{j=1}^{i-1} \sin \gamma_j \cos \gamma_j$, $i = 1, \ldots, nm-q-1$, $z_{nm-q}^+ = R^* \prod_{j=1}^{nm-q-1} \sin \gamma_j$. Then

$$I \geq J_m f_m k_1 \epsilon_1^a \epsilon_2^{\frac{1}{2}n(n+1)} k_2/\lambda_M^k \int_{\epsilon_4}^{\epsilon_3} R^{*nm-q-k-1} dR^{*2} \int_0^{\gamma_0} [(z_0'z_0)^{\frac{1}{2}} \cos \gamma_1 - v_0]^k \sin \gamma_1^{nm-q-2} d\gamma_1$$

where k_2 is the constant of integration obtained by integrating with respect to γ_i , $i = 2, \ldots, nm-q-1$, and γ_0 is defined by $(z_0'z_0)^{1/2} \cos \gamma_0 = v_0$. Clearly this tends to ∞ as $\epsilon_4 \to 0$ if $nm-q-k \leq 0$, or if $k \geq nm-q$.

<u>End of Proof</u>. Note that the assumption that $\Psi(p,w)$ is locally positive definite is required to ensure that the local parameterization is as stated. It may also play an important part in our subsequent discussions of assumption (E). (E) also requires that $(h'\eta_0)$ is non zero, and this is an important requirement for the choice of (p_0, w_0) . Indeed the use of the theorem requires now that we discuss in detail whether it is possible to find a point (p_0, w_0) where all the assumptions (A) to (E) are satisfied. The conditions (A) to (D) are reasonably general, but (E) is rather complex. In addition to $(h'\eta_0) \neq 0$, we require $H'(\theta_0)Z_p'(\Psi_0^{1/2} \otimes I)d_0 \neq 0$. The elements of the vector $H'(\theta_0)Z_p'(\Psi_0^{1/2} \otimes I)d_0$ can be regarded as functions of θ_0 , and the equation of the elements of this vector to zero define a differential manifold. It is only if this manifold contains all the points of the zero-set at which $h'\eta_0 \neq 0$ that we will find it impossible to

satisfy condition (E). However further discussion requires detailed consideration of the parameterization of the zero-set.

3. The 3SLS Parameterization of the Zero-Set

Since $z(p^*(\theta_1), w) = 0$ for all θ_1, differentiating at θ_0 we have $Z_p(\partial p^*/\partial \theta_1) = 0$, so that Z_p has rank less than or equal to $nm-q$. Now write

$$
\begin{aligned}
E(p) &= T^{-1/2}(\Psi^{1/2} \otimes (Z'Z)^{-1/2}(Z'X))K \\
&= (\Psi^{1/2} \otimes [\hat{P}' : M^{1/2}])K \\
&= (\Psi^{1/2} \otimes \hat{P}')[I \otimes (I : 0)]K + (\Psi^{1/2} \otimes I)[I \otimes (0 : M^{1/2})]K \\
&= (\Psi^{1/2} \otimes \hat{P}')K_1 + (\Psi^{1/2} \otimes I)K_2 ,
\end{aligned} \tag{8}
$$

where the last equation defines K_1 and K_2. Note that $E'E = K'(\Psi \otimes R)K$, and taking λ to be the smallest root of $E'E$ we differentiate $(E'E - \lambda I)\eta^* = 0$. The result is

$$\left(\frac{\partial}{\partial p}(E'E) - \frac{\partial \lambda}{\partial p} I\right)\eta^* + (E'E - \lambda I)\frac{\partial \eta^*}{\partial p} = 0 .$$

Note that since L was defined as an orthogonal matrix $\eta^{*\prime}\eta^* = 1$, so that $\eta^{*\prime}(\partial \eta^*/\partial p) = 0$. Thus

$$\frac{\partial \lambda}{\partial p} = \eta^{*\prime}\,\frac{\partial(E'E)}{\partial p}\,\eta^* ,$$

and

$$\frac{\partial \eta^*}{\partial p} = -(E'E - \lambda I)^{-}\,\frac{\partial(E'E)}{\partial p}\,\eta^* ,$$

provided λ is a single root, and $(E'E - \lambda I)^-$ is the usual generalized inverse. If $\lambda = 0$ so that $E\eta^* = 0$, then $\partial\lambda/\partial p = 0$ and

$$\frac{\partial\eta^*}{\partial p} = -(E'E)^- E' \frac{\partial(E\eta_0)}{\partial p} ,$$

where η_0 is treated as a constant when differentiating. This formulation has the advantage that $\partial(E\eta_0)/\partial p$ can be thought of as a square matrix. Then

$$\text{(9)} \qquad \partial z/\partial p = \partial(E\eta^*)/\partial p = (I - E(E'E)^- E')\partial(E\eta_0)/\partial p .$$

The first factor is an idempotent matrix of rank $nm - N + 1$, orthogonal to the matrix E. Now considering the equations

$$\text{(10)} \qquad E(p)\eta_0 = 0 ,$$

for arbitrary η_0, clearly a sufficient condition that p belongs to the zero set is that the equations are satisfied for some η_0, and there are as many equations as there are elements of the p vector. Thus it may be possible to solve the equations for p uniquely. Since the equations are linear in p, there may be no, one, or an infinity of solutions. When there is a unique solution there will be a neighborhood of this η_0 in which there is a parameterization of the zero-set with p a function of η_0. Since this function is zero order homogeneous, it is possible to make one element of η_0 equal to one, and to regard the remaining elements of η_0 as forming an N-1 vector of parameters θ_1. All points at which such a parameterization is valid form an (N-1) dimensional manifold which will be called the full rank manifold. Indeed using the form of equation (8) we can obviously transform these equations to the form

(11) $$\Psi^{1/2}\Gamma_1\hat{P} + \Psi^{1/2}\Gamma_2 = 0$$

where $\text{vec}\ \Gamma_1 = K_1\eta_0$ and $\text{vec}\ \Gamma_2 = K_2\eta_0$. A necessary and sufficient condition that $\hat{P}$ is uniquely determined by (8) is that Ψ is non-singular and Γ_1 is non-singular. Only vectors η_0 which give a non-singular Γ_1 uniquely determine a point of the full rank manifold.

For such points differentiating $E(p)\eta^* = 0$ we find that

$$\left(\frac{\partial(E\eta_0)}{\partial p}\right)\frac{\partial p^*}{\partial\theta} + E^*(p) = 0 ,$$

where $E^*(p)$ consists of the (N-1) columns of $E(p)$ which correspond to the elements of θ . In accordance with condition (B) of the theorem we assume (p_0, w_0) to be chosen so that $E^*(p)$ is of rank N-1 .

Then defining H^* so that $H^*H^{*\prime} = I - E(E'E)^-E'$, and $H^{*\prime}H^* = I$, and defining $\Phi^* = H^{*\prime}(\partial(E\eta_0)/\partial p)H'$, we have

$$Z_p = (I - E(E'E)^-E')\ \frac{\partial(E\eta_0)}{\partial p} = H^*H^{*\prime}\ \frac{\partial(E\eta_0)}{\partial p} = H^*\Phi^*H ,$$

and
$$\Phi^*\Phi^{*\prime} = H^{*\prime}\ \frac{\partial(E\eta_0)}{\partial p}\ Q(\theta_0)\left(\frac{\partial(E\eta_0)}{\partial p}\right)' H^*$$

$$= H^{*\prime}(\Psi_0^{1/2}\Gamma_1\Gamma_1'\Psi_0^{1/2} \otimes I)H^* ,$$

since
$$H^{*\prime}\ \frac{\partial(E\eta_0)}{\partial p}\left(\frac{\partial p^*}{\partial\theta_1}\right) = H^{*\prime}E^* = 0 .$$

Thus Φ^* is non-singular if Γ_1 is non-singular. Consider now the condition that $H'(\partial(E\eta_0)/\partial p)'H^*H^{*\prime}[\Psi_0^{1/2} \otimes I]d_0 \neq 0$, which is required for (E) of Theorem 1. This latter vector equals $\Phi^{*\prime}H^{*\prime}(\Psi_0^{1/2} \otimes I)d_0$, so that from the non-singularity of Φ^*, the condition is equivalent to $H^{*\prime}(\Psi_0^{1/2} \otimes I)\text{vec}(BP_0 + CM^{1/2}) \neq 0$. Now since H^* is the annihilator of E^* this is equivalent to saying that there exists no vector η^+ such that

$$(\Psi_0^{1/2} \otimes I)\text{vec}(BP_0 + CM^{1/2}) + E^*\eta^+ = 0 .$$

Now by adding an extra element to η^+, which is zero, in the position, which was previously standardized as one, we can write $E^*\eta^+ = E(p_0)\tilde{\eta}$. ($\tilde{\eta}$ has all its elements the same as η^+ except for the extra zero.)

Then writing $E(p_0) = (\Psi_0^{1/2} \otimes P_0')K_1 + (\Psi_0^{1/2} \otimes I)K_2$ we have that there should be no $\tilde{\eta}$ such that

$$\text{vec}(BP_0 + CM^{1/2}) + (I \otimes P_0')K_1\tilde{\eta} + K_2\tilde{\eta} = 0 . \tag{12}$$

Now it is not impossible given that (12) are fairly complex conditions that there are values of P_0 such that P_0 is a point of the zero-set and so satisfies (10), and such that a vector $\tilde{\eta}$ can be found which satisfies (12), but it seems unlikely that this would be true for all points of the full-rank manifold. This is, however, a weak argument, and we can bolster it a little by considering points in the neighborhood of a point p_0 which satisfies both (10) and (12). We wish to establish that a further set of equations must hold at this point if there is not to be at least one point in the neighborhood at which (10) is satisfied but not (12).

The argument can be formulated as follows. Write equations (12) as a vector of equations of the form $g(p, \tilde{\eta}) = 0$, and equation (10) as the vector of equations $g^*(p, \theta_1) = 0$, where θ_1 are the non-standardized elements of η^*. Then we assume that $p = p_0$, $\theta_1 = \theta_0$, $\tilde{\eta} = \tilde{\eta}_0$, satisfy both sets of equations. Consider the differential equations of the form

$$(\partial g/\partial p)dp + (\partial g/\partial\tilde{\eta})d\tilde{\eta} = 0$$

$$(\partial g^*/\partial p)dp + (\partial g^*/\partial\theta_1)d\theta_1 = 0 .$$

Assuming that p_0 is a point of the full rank manifold, we have that $\partial g^*/\partial p$ is non-singular so that if for any $d\theta_1$, we have the first equation satisfied we can eliminate dp to obtain

$$-[\partial g/\partial p(\partial g^*/\partial p)^{-1}(\partial g^*/\partial\theta_1)]d\theta_1 + (\partial g/\partial\tilde{\eta})d\tilde{\eta} = 0 .$$

Writing $G^* = -(\partial g/\partial p)(\partial g^*/\partial p)^{-1}(\partial g^*/\partial\theta_1)$, it follows by considering a change in θ_1 only one of whose elements is non-zero, that we must have that every column of G^* is a linear combination of the columns of $\partial g/\partial\tilde{\eta}$, or in other words that the matrix $(G^* : \partial g/\partial\tilde{\eta})$ is of rank less than or equal to N-1 . This is equivalent to requiring that the matrix

$$\begin{pmatrix} 0 & : \partial g/\partial p & : \partial g/\partial\tilde{\eta} \\ \partial g^*/\partial\theta_1 & : \partial g^*/\partial p & : 0 \end{pmatrix}$$

is of rank $\leq N-1+nm$. However specifically in this case we have that $\partial g/\partial\tilde{\eta} = \partial g^*/\partial\theta_1 = E^*$, and $\partial g^*/\partial p = \Gamma_1 \otimes I$, and $\partial g/\partial p = (B + \tilde{\Gamma}_1) \otimes I$, where $\text{vec}(\tilde{\Gamma}_1) = K_1\tilde{\eta}$, so that writing $G^+ = -\partial g/\partial p(\partial g^*/\partial p)^{-1} =$ $= -(B + \tilde{\Gamma}_1)\Gamma_1^{-1} \otimes I$, we require that $(G^+E^* : E^*)$ is of rank $N-1$. This is equivalent to

$$G^+E^* = E^*\Lambda \ , \tag{13}$$

for some $(N-1) \times (N-1)$ matrix Λ, and this requires that E^* is some $(N-1) \times (N-1)$ linear combination of the $N-1$ latent vectors of G^+. Indeed in this case writing $G = -(B + \tilde{\Gamma}_1)\Gamma_1^{-1}$ we require that $(G \otimes I)E^* = E^*\Lambda$, and writing the latent vectors of G as H^+, we must have

$$E^* = (H^+ \otimes I)\begin{pmatrix} D_1 & & & 0 \\ & D_2 & & \\ & & \ddots & \\ 0 & & & D_n \end{pmatrix} E^+$$

where E^+ is a square $(N-1) \times (N-1)$ matrix, and each D_i has m rows but may have any number of columns provided the total number of columns is $N-1$. This set of conditions is quite formidable, and makes it clear that it will be usually possible to discover some point of the zero-set which has a non-zero z_0 in the neighborhood of any point for which $z_0 = 0$.

However not every point of the zero-set is a point of the full rank manifold, and it is necessary to consider other points of the zero-set. Such points are characterized by $\det \Gamma_1 = 0$. If for some vector η^* Γ_1 is of rank $n^* < n$, then as η^* tends to this value the corresponding p satisfying $E(p)\eta^* = 0$ tends to infinity unless $(\Gamma_1 : \Gamma_2)$ is of rank n^* for this η^*. Thus if p remains finite as η^* tends to its limit there must exist an $(n-n^*) \times n$ full rank matrix β such that $\beta\Gamma_1 = 0$, $\beta\Gamma_2 = 0$ so that the n equations (10) are equivalent to only n^* equations. It then follows that $m(n-n^*)$ elements of p can be chosen arbitrarily, in finding a point of the zero-set for this value of π^*, and the zero-set contains an affine linear subspace. The condition that $(\Gamma_1 : \Gamma_2)$ is of rank n^* can be represented in the form that $(n-n^*)(m+n-n^*)$ submatrices of dimension $(n^*+1) \times (n^*+1)$ should be singular: each condition equates a polynomial of degree (n^*+1) in π^* to zero. Thus if $(n-n^*)(m+n-n^*) < N-1$ there may be a set of points in π^* space forming a differential cone of dimension $N-(n-n^*)(m+n-n^*)$ in π^* space. Since the $m(n-n^*)$ elements of P_0 can be chosen arbitrarily, the resulting differential manifold in p space has in general dimension $N-(n-n^*)^2-1$. It follows that the full rank manifold usually provides a lesser upper bound than this restricted rank manifold.

However if the K matrix has a special form this may not be true. This certainly happens in the following special case.

Suppose that, for some constant matrices G_i, $i = 1, \ldots, 5$,

(14) $$K_1 G_1 = (G_2 \otimes G_3) G_4$$

(15) $$K_2 G_1 = (G_2 \otimes I_m) G_5 .$$

G_1 is $N \times N_1$, G_2 is $n \times N_2$ of rank N_2, G_3 is $n \times N_3$ of rank N_3, G_4 is $(N_2 N_3) \times N_1$, G_5 is $(N_2 m) \times N_1$. Then if $\pi^* = G_1 \theta^*$, where some element of θ^* is standardized to be one, then $E(p)\pi^* = 0$ is equivalent to

(16) $$(G_2 \otimes \hat{P}' G_3) G_4 \theta^* + (G_2 \otimes I_m) G_5 \theta^* = 0 .$$

Since G_2 is of rank N_2 this is equivalent to

(17) $$(I_{N_2} \otimes \hat{P}'G_3)G_4\theta^* + G_5\theta^* = 0 \ .$$

Provided $m(N_2 - N_3) \leq N_1 - 1$, it follows that this may determine a differential manifold in p space of dimension $(n - N_2)m + N_1 - 1$. If this is greater than $N-1$, a suitable choice of p_0 on this manifold would put an upper bound on the maximal exponent $N_2m - N_1 + 1$. This example is sufficient to show that the determination of the maximal exponent depends in a rather complex way upon the exact formulation of the model.

A much more important special form of model arises when the constraints are separable in the sense that each constraint refers only to the coefficients of a single equation, so that the constraints can be written

(18) $$\Phi_i a_i = \phi_{0i} \ , \quad i = 1, \ldots, n \ .$$

Note that for identification we must have that $\phi_{0i} \neq 0$, all i.

For each i we have a set of r_i constraints containing only the coefficients of the i^{th} equation. We write the corresponding K matrix in the form

$$K = \begin{pmatrix} K_1^* & 0 & \cdots & 0 \\ 0 & K_2^* & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & K_n^* \end{pmatrix}$$

where K_i^* is $(n+m) \times (n+m-r_i)$.

Partitioning η^* in the same way we have

(19) $$(\hat{P}' : M^{1/2})K_i^*\eta_i^* = 0 \, , \quad i = 1, \ldots, n \, .$$

Writing $K_i^* = \begin{pmatrix} K_{i1} \\ M^{-1/2}K_{i2} \end{pmatrix}$, $E_i = \hat{P}'K_{i1} + K_{i2}$, and we have (19) in the form

(20) $$(\hat{P}'K_{i1} + K_{i2})\eta_i^* = 0 \, .$$

Note that if (20) is satisfied for some η_i^* which is non-zero, we can put $\eta_j^* = 0$ for $j \neq i$. Thus in this case the zero-set, given by $\psi(p_0, w_0) = 0$, consists of the union of all $\hat{P}$ which satisfy (20) for some i.

There is a very special case where for some i and j $K_{i1} = K_{j1}C^*$, $K_{i2} = K_{j2}C^*$, for some matrix C^*, where the set defined by (20) for the i^{th} equation is contained in the set for the j^{th} equation. This makes it difficult to apply Theorem 1 to the i^{th} equation, as will become apparent. However this is a very special case and it will not be discussed further.

We assume then that it is always possible to find a point p_0 of the zero-set belonging to the i^{th} equation, and so satisfying (20) for i, which does not belong to any other equation, so that η_0 has a subvector $\eta_{i0} \neq 0$, but $\eta_{j0} = 0$ for all $j \neq i$. Suppose that we consider in Theorem 1 a vector h which has all its elements zero except those corresponding to parameters of the j^{th} equation. Then $h'\eta_0 = 0$ unless $j = i$, and it follows that p_0 can only be used to discuss the moments of the parameters of the i^{th} equation.

Now writing $K_{i1} = \begin{pmatrix} \kappa'_{i1} \\ \kappa^*_i \end{pmatrix}$, where κ_{i1} is the first row of K_{i1}, and reordering if necessary the columns of $\hat{P}$ so that $\kappa_0 = \kappa'_{i1}\eta_{i0} \neq 0$ (η_{i0} is the subvector of η_0 corresponding to equation i), and writing $\hat{P} = \begin{pmatrix} \hat{p}' \\ P^+ \end{pmatrix}$, we have

$$\hat{p} = (K_{12}\eta^*_i - P^{+'}\kappa^*_i\eta^*_i)/(\kappa'_{i1}\eta^*_i) \ .$$

$\hat{P}$ can thus be represented as depending on the $m(n-1)$ elements of P^+, and of the $N_i - 1$ elements of η^*_i (where one element of η^*_i is standardized to be one). We then have as $\partial p^*/\partial\theta$ at $p = p_0$

$$\partial p^*/\partial\theta = \begin{pmatrix} E^*_i/\kappa_0 & -\eta^{*'}_i\kappa^{*'}_i/\kappa_0 \otimes I \\ 0 & I_{(n-1)m} \end{pmatrix}$$

where E^*_i is the $m \times (N_i - 1)$ obtained by taking the $(N_i - 1)$ columns of $(\hat{P}'K_{i1} + K_{i2})$ which correspond to non-standardized elements of η^*_i.

Thus $H(\theta)$ being orthogonal to $\partial p^*/\partial\theta$ can be written

$$H(\theta) = \frac{1}{\kappa^*}(K_{i1}\eta^*_i) \otimes H_i$$

where $\kappa^{*2} = (\eta^{*'}_i K'_{i1}K_{i1}\eta^*_i)$. and H_i satisfies $H'_i E^*_i = 0$ and $H_i H'_i = I_{n-N_i-1}$. It is convenient now to reorder the equations of the model so that we are considering the first equation, and then to change the definition of $\Phi_0^{1/2}$. Up to now we have been assuming that $\Psi_0^{1/2}$ is symmetric, but it is now convenient to assume $\Psi_0^{1/2}$ is upper triangular so that now we have $(\Phi_0^{1/2})'\Phi_0^{1/2} = \Phi_0$.

Then since

$$E^* = (\Phi_0^{1/2} \otimes I)\begin{pmatrix} E_1^* & 0 & \cdots & 0 \\ 0 & E_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & E_n \end{pmatrix} ,$$

$$H^{*'} = \begin{pmatrix} H_1' & 0 & \cdots & 0 \\ 0 & H_2^{*'} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & H_n^{*'} \end{pmatrix}(\Phi_0^{-1/2} \otimes I)$$

where $H_1'E_1^* = 0$, and $H_r^{*'}E_r = 0$, $r = 2, \ldots, n$. Also $E\eta_0$

$$E\eta_0 = \psi_{11}\begin{pmatrix} E_1\eta_{10} \\ 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

where ψ_{11} is the one-one element of $\Phi_0^{1/2}$ so that

$$\partial(E\eta_0)/\partial p = \psi_{11}\begin{pmatrix} (\eta_{10}'K_{11}') \otimes I \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and

$$(\partial(E\eta_0)/\partial p)H = \psi_{11}\kappa^* \begin{pmatrix} H_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

Finally,

$$\Phi^* = H^{*\prime} \frac{\partial(E\eta_0)}{\partial p} H = (\psi''\psi_{11})\kappa^* \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

where ψ'' is the one-one element of $\Phi_0^{-1/2}$

$$= \kappa^* \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

since $\psi''\psi_{11} = 1$. Thus

$$\begin{aligned} \Phi^{*\prime}H^{*\prime}(\Psi_0^{1/2} \times I)d_0 &= \kappa^*(H_1', 0, 0, \ldots, 0)\mathrm{vec}(BP_0 + CM^{1/2}) \\ &= \kappa^* H_1'(P_0'b_1 + M^{1/2}c_1) . \end{aligned} \tag{20}$$

Now from its definition $\kappa^* \geq |\kappa_0|$, and we assumed that we chose to order the endogenous variables so that $|\kappa_0|$ is non-zero, which ensures that κ^* is also non-zero.

Thus (20) is zero, if and only if

$$P_0'b_1 + M^{1/2}c_1 = E_1^*\eta_1^+ \tag{21}$$

for some vector η_1^+ .

Now define $\tilde{\eta}_1$ to be a vector of dimension N_1 such that it is the same as η_1^+ except for an extra zero in place of the standardized element of η_1^* .

Then (21) can be written in the form

$$P_0'b_1 + M^{1/2}c_1 = (P_0'K_{11} + K_{12})\tilde{\eta}_1 \ . \tag{22}$$

We must now consider whether for every P_0 satisfying

$$(P_0'K_{11} + K_{12})\eta_{10} = 0 \tag{23}$$

for some η_{10} , there exists an $\tilde{\eta}_1$ such that

$$P_0'b_1 + M^{1/2}c_1 = (P_0'K_{11} + K_{12})\tilde{\eta}_1 \ .$$

We will show that if we start with a P_0 such that both (22) and (23) are satisfied for suitable η_{10} and $\tilde{\eta}_1$, it is always possible to find a P_0^* close to this P_0 , such that not both (22) and (23) are satisfied.

For suppose on the contrary that for some P_0 (22) and (23) are satisfied, and for all $\Delta\eta_{10}$, and corresponding ΔP_0 , it is possible to find $\Delta\tilde{\eta}_1$ satisfying (22). Then we must have that, if

$$\text{(24)} \qquad \Delta P_0'(K_{11}\eta_{10}) + E_1^*\Delta\eta_1^* = 0$$

for all possible matrices ΔP_0, and vectors $\Delta\eta_1^*$, where $\Delta\eta_1^*$ has elements corresponding only to the variable elements of η_1^*, then

$$\text{(25)} \qquad \Delta P_0'(K_{11}\tilde{\eta}_1 - b_1) + E_1^*\Delta\eta_1^+ = 0$$

for some $\Delta\eta^+$.

(24) is equivalent to $H_1'\Delta P_0'(K_{11}\eta_{10}) = 0$, and (25) to

$$H_1'\Delta P_0'(K_{11}\tilde{\eta}_1 - b_1) = 0 .$$

We require to discuss under what circumstances we can choose ΔP_0 so that

$$(\eta_{10}'K_{11}' \times H_1')\text{vec } \Delta P_0 = 0$$

and

$$(\tilde{\eta}_1'K_{11}' - b_1') \times H_1' \text{ vec } \Delta P_0 \neq 0 .$$

Given that the column of H_1 are linearly independent it is clear that these last two sets of equations can be satisfied if

$$\tilde{\eta}_1'K_{11}' - b_1' \neq \lambda\eta_{10}'K_{11}'$$

for some scalar λ. Suppose on the contrary that

$$b_1 = K_{11}(\tilde{\eta}_1 - \lambda\eta_{10}) .$$

Then $M^{1/2}c_1 - K_{12}(\tilde{\eta}_1 - \lambda\eta_{10}) = -P_0'(b_1 - K_{11}(\tilde{\eta}_1 - \lambda\eta_{10})) = 0$ from (22) and (23).

These equations can then be combined in the form $a_1 = K_1^*(\tilde{\pi}_1 - \lambda\pi_{10})$. But then

$$\Phi_1 a_1 = \Phi_1 K_1^*(\tilde{\pi}_1 - \lambda\pi_{10}) = 0$$

since $\Phi_1 K_1^* = 0$, from the conditions that the K_1^* yield a valid parameterization of the constraints.

But this contradicts the condition (18) that $\Phi_1 a_1 = \phi_{01} \neq 0$, as a condition for identification of the first equation.

So we conclude that $\tilde{\pi}_1' K_{11}' - b_1'$ cannot be proportional to $\pi_{10}' K_{11}'$, so that there is a point in the neighborhood of P_0 at which either (22) or (23) is not satisfied. Thus a suitable point in the zero set for the applications of Theorem 1 can always be found.

4. A Lower Bound for the Maximal Exponent

The next theorem establishes sufficient conditions for the existence of moments of a given order. To simplify the discussion use is made of inequalities on the function whose moment is required. From the definition of the 3SLS estimator as the constrained minimum of $\mathrm{tr}(\Psi ARA')$ we have

$$\mathrm{tr}(\Psi\hat{A}R\hat{A}') \leq \mathrm{tr}(\Psi\bar{A}R\bar{A}') \tag{26}$$

where we temporarily use $\bar{A}$ to mean the true value of A . Thus

$$\mathrm{tr}(\Psi(\hat{A}-\bar{A})R(\hat{A}-\bar{A})') \leq -2\ \mathrm{tr}(\Psi\bar{A}R(\hat{A}-\bar{A})') ,$$

so that using the Cauchy inequality

$$\mathrm{tr}(\Psi(\hat{A}-\bar{A})R(\hat{A}-\bar{A})') \leq 4\ \mathrm{tr}(\Psi\bar{A}R\bar{A}') . \tag{27}$$

Now we continue to assume that if Ψ is rescaled so that its largest latent root is one, then the smallest root of the rescaled matrix, which will be denoted by μ, is bounded above zero. (In the last section we only made this assumption for some neighborhood of p_0, w_0.) Then

$$\mathrm{tr}((\hat{A}-\bar{A})R(\hat{A}-\bar{A})') \leq 4\ \mathrm{tr}(\bar{A}R\bar{A}')/\mu ,$$

or

$$(\hat{\alpha}-\alpha)'(K'(I \otimes R)K)(\hat{\alpha}-\alpha) \leq 4\ \mathrm{tr}(\bar{A}R\bar{A}')/\mu .$$

Now let λ_m be the smallest root of $K'(I \otimes R)K$, so that

$$T(\hat{\alpha}-\alpha)'(\hat{\alpha}-\alpha) \leq 4T\ \mathrm{tr}(\bar{A}R\bar{A}')/\lambda_m \mu .$$

Note that $T\ \mathrm{tr}(\bar{A}R\bar{A}') = \mathrm{tr}((\bar{A}X'Z)(Z'Z)^{-1}(Z'X\bar{A}')) = \mathrm{tr}((U'Z)(Z'Z)^{-1}(Z'U)) = O(1)$ as $T \to \infty$. However as a further simplification we have that

$$\lambda_m \leq \det(K'(I \otimes R)K)/[\mathrm{tr}(K'(I \otimes R)K)]^{n-1} ,$$

since $K'(I \times R)K$ is an $n x n$ non-negative definite matrix. Thus we can write

$$T(\hat{\alpha}-\alpha)'(\hat{\alpha}-\alpha) \leq \phi(p)/\psi^*(p) , \tag{28}$$

where $\psi^*(p) = \det(K'(I \otimes R)K)$ and $\phi(p) = 4T\ \mathrm{tr}(ARA')[\mathrm{tr}(K'(I \otimes R)K)]^{n-1}/\mu$. Note that both $\phi(p)$ and $\psi^*(p)$ are polynomials in p.

Suppose we now wish to consider the $(\frac{1}{2}k)^{th}$ moment of $T(\hat{\alpha}-\alpha)'(\hat{\alpha}-\alpha)$. We can obviously consider the same moment of $\phi(p)/\psi^*(p)$, knowing that if the latter exists then so does the former. Considering then the latter moment, we can generalize as follows:

Suppose:

A^*. $\sqrt{T}(p - E(p))$ is a stochastic vector with finite absolute moments up to exponent $N_0 > Nk$, which are uniformly bounded as $T \to \infty$, and that the distribution of p is absolutely continuous with density $f(p)$.

B^*. $\phi(p)$ is a polynomial of degree $2N$ and of $O(1)$ as $T \to \infty$, in the sense that if it is expressed in the form $\phi(p) = \phi_d(\sqrt{T}(p - E(p)))$ all the coefficients of ϕ_d are bounded as $T \to \infty$.

Now define $I_T^* = \int|\phi(p)|^{1/2\,k} f(p)dp$, and note that the conditions on the moments of $f(p)$ ensure that I_T^* is bounded as $T \to \infty$, and define $f^*(p) = |\phi(p)|^{1/2\,k} f(p)/I_T^*$. Then $f^*(p)$ is a probability density with moments up to exponent $N_0 - Nk$ uniformly bounded as $T \to \infty$. We consider moments of the form $\int|\psi^*(p)|^{-1/2\,k} f^*(p)dp$. Clearly if such a moment is finite and bounded as $T \to \infty$ then the $(\frac{1}{2}k)^{th}$ moment of $T(\hat{\alpha}-\alpha)'(\hat{\alpha}-\alpha)$ has the same properties. The next theorem considers sufficient conditions for the existence of moments of $1/\psi^*(p)$. Note as an important special case that if $f(p)$ has uniformly bounded moments of $\sqrt{T}(p - E(p))$ of all orders then so has $f^*(p)$.

<u>Lemma</u>. If $1 - F(g^*) = P(|g| \geq g^*) \leq c/g^{*r}$, all $g^* > g_0$ and some $c > 0$ independently of T then $E(|g|^s)$ exists and is bounded for all T, if $s < r$.

The proof is by integration by parts.

<u>Definition</u>. A non-negative definite function $\psi^*(p)$ of a $q \times 1$ vector p will be said to be inverse moment regular on a closed connected set with non-empty interior $\mathcal{P}^*$, if it can be written in the form

(29) $$\psi^*(p) = \sum_{i=1}^{k_0} \lambda_i(p)|\xi_{i0}(p)|^{r_i} R_i \ ,$$

where either $R_i = 1$, or $R_i = \sum_{j=1}^{k_i} \xi_{ij}^2(p)$, provided that $r_i k_i > 1$ in the latter case, and that $k^+ = \sum_{j=0}^{k_0} k_i \leq q$, where we define $k_i = 0$ if $R_i = 1$. For all $p \in \Omega^*$ we require $\lambda_i(p) \geq b > 0$, for some b , and that the function ξ_{ij} , $j = 1, \ldots, k_i$, $i = 1, \ldots, k_0$, have continuous first derivatives such that $\partial\xi/\partial p$ is strictly of rank k^+ everywhere in Ω^* .

Note that (29) is equivalent to $\psi^*(p)/b \geq \sum_{i=1}^{k_0} |\xi_{i0}|^{r_i} R_i$.

<u>Theorem 2</u>. If (A^+) the set Ω^+ defined by $\psi^*(p) \leq \epsilon^+ > 0$ is non-empty and such that $\psi^*(p)$ is inverse moment regular throughout Ω^+ , (B^+) p is a stochastic vector such that for any set of elements $\tilde{p}$ forming a vector of order $q^* \leq k^+$, if p^* is the complement vector of $q - q^*$ element, then the conditional density $f(\tilde{p}/p^*)$ is bounded for all p in Ω^+ : then $E(|\psi^*(p)|^{-1/2\, k}) < \infty$ if $k < 2 \sum_i (1/r_i)$.

<u>Proof</u>. We make use of the lemma and so consider the probability $P(\psi^*(p) \leq \psi_0)$ for $0 \leq \psi_0 \leq \epsilon^+$. We initially partition the set S^* defined by $\psi^*(p) \leq \psi_0$ into 2^{k^*} subsets, where k^* is the number of terms in (29) for which $R_i \neq 1$. Suppose these terms number $i = 1, \ldots, k^*$. For each such i distinguish between the points of S^* for which $R_i < 1$, and those for which $R_i \geq 1$, for every $i = 1, \ldots, k^*$. Thus in each subset of S^* , there is a certain set of integers $i \leq k^*$, such that

$R_i < 1$ for all p in the subset. In each subset we divide up the p vector into $(\tilde{p}, p^*)$ such that $\tilde{p}$ has a number of elements equal to k_0 plus the sum of the k_i for the set of i such that $R_i < 1$. In each subset we transform the variables of integration from p to the set of variables made up of p^*, π_{i0}, $i = 1, \ldots, k_0$, and π_{ij}, $j = 1, \ldots, k_i$, for i such that $R_i < 1$. This set of π will be denoted by π^* later. We now consider the imposition of a lower bound on the Jacobian of the transformation.

Suppose D_s, $s = 1, \ldots, C^q_{k^+}$, denotes the $k^+ \times k^+$ submatrices of $\partial\pi/\partial p$ arranged in some order. Define $(\lambda^*(p))^2 = \max_s$(smallest latent root of $D_s'D_s$). By the assumption that $\partial\pi/\partial p$ is strictly of rank k^+, we mean that $\lambda^*(p)$ is non-zero everywhere on Ω^*, and by continuity if Ω^* is compact we can then find $\lambda_0 > 0$, such that $\lambda^*(p) \geq \lambda_0$. In fact if Ω^* is unbounded, then we reinterpret "strict" above to mean $\lambda^*(p) \geq \lambda_0 > 0$ for some λ_0. For each p denote the s in the definition of $\lambda^*(p)$ by $s(p)$. We now partition each subset defined above into a finite set of not more than $C^q_{k^+}$ subsets such that each of these sub-subsets has the smallest root of $D_{s^*} > \lambda_0$ for some particular s^*. Denote $D_{s^*} = D^*$ and partition $D^*_s = D^*$, and partition D^* into $\begin{pmatrix} D^+ \\ D^{++} \end{pmatrix}$, where D^+ are the rows of D^* which correspond the particular π_{ij} listed above for this original subset. Then $D^{+'}D^+$ is a principal minor of $D^{*'}D^*$, so that its smallest latent root is greater than $\lambda^* > \lambda_0$. Thus if n^* is the number of elements in $\tilde{p}$ $\det(D^{+'}D^+) \geq \lambda_0^{n^*}$. Then $\det(D^{+'}D^+) = \sum_s J_s^{*2}$, where J^*_s is the determinant of some $n^* \times n^*$ submatrix of D^+ , and the summation is over all such submatrices. If $J_M = \max|J_k^*|$, then

$$J_M^2 \geq \lambda_0^{n^*}/C_{n^*}^{k^+} \geq B^2 \ ,$$

where the last equation defines B . This shows that indeed we can partition our original subsets into not more than $C_{n^*}^q$ subsets such that for some particular choice of $(\tilde{p}, p^*)$ the Jacobian $|J| \geq B$, where this is the Jacobian of the transformation from p to (p^*, ξ^*) , and the ξ^* is the set of ξ_{ij} originally used to define the original subset of which this is a sub-subset. Denote the sub-subsets by ω_L , and then we have

$$(30) \qquad P(\psi^*(p) \leq \psi_0) = \sum_L \int_{\omega_L : \psi(p) \leq \lambda_0} \frac{f(\tilde{p}|p^*) d\xi^*}{J} dF(p^*)$$

where $F(p^*)$ is the c.d.f. for p^* , and $d\xi^*$ is the differential element of the set of ξ_{ij} denoted by ξ^* earlier, i.e. that listed in the first paragraph of the proof. But if $\psi(p) \leq \psi_0$, $|\xi_{i0}| \leq (\psi_0/bR_i)^{1/r_i}$, so that assuming $f(\tilde{p}|p^*) < f^*$, and $1/J \leq 1/B$, we have by integrating with respect to ξ_{i0} , $i = 1, \ldots, k_0$ that $P(\psi^*(p) \leq \psi_0)$
$\leq (f^*/B)(\psi_0/b)^{\Sigma 1/r_i} C_1^* \sum_L \int_{\omega_L} \prod_{i=1}^{k_0} R_i^{-1/r_i} d\xi^+ dF(p^*)$, where C^* is a constant of integration, and $d\xi^+$ is the differential element of those ξ_{ij} in ξ^* for which $j > 0$. Note that for those i for which $R_i \geq 1$ in ω_L , we can replace R_i by one in the above inequality. For the R_i such that $R_i < 1$ in ω_L , we can integrate w.r.t. ξ_{ij} using a polar coordinate transformation, so as to have R_i as the radius of the hypersphere. As a result we obtain

$$P(\psi^*(p) \leq \psi_0) \leq C_2^*(f^*/B)(\psi_0/b)^{\Sigma 1/r_i} \sum_L \int_{\omega_L} \Pi_i R_i^{k_i - 1 - 1/r_i} dR_i dF(p^*) .$$

This last integral obviously converges provided $k_i - 1/r_i > 0$. Then application of the lemma gives the required result.

<u>End of Proof</u>. This proof is not adequate to show that the moment is uniformly bounded as $T \to \infty$, since in general $f(\tilde{p}|p^*)$ is $O(T^{n^*/2})$. Thus we need stronger assumptions.

Uniform bounding will be discussed only in the case where the errors are normally distributed since any formally more general assumptions do not have much intuitive content. When p is normally distributed it is convenient to redefine p so that $\sqrt{T}(p - \bar{p}) \sim N(0,I)$. In the 3SLS case we define $\tilde{P} = \Omega_v^{-1/2}\hat{P}$, and then define $p = \text{vec}\ \tilde{P}$. Our previous $\psi^*(p)$ can be easily expressed as a function of these new variables.

We then have the following modification of Theorem 2.

<u>Corollary</u>. If in place of (B^+) we assume $\sqrt{T}(p - \bar{p}) \sim N(0,I)$, and $\psi^*(p) > 0$, then the moments for $k < 2 \sum_i (1/r_i)$ are uniformly bounded as $T \to \infty$.

<u>Proof</u>. Choose $\epsilon^+ < \psi^*(\bar{p})$. Then define $\delta_0^2 = \inf \|p - \bar{p}\|^2 : \psi^*(p) \leq \epsilon^+$. Then in this case writing $z = \sqrt{T}(p - \bar{p})$, and partitioning z in the same way that $p = (\tilde{p}, p^*)$ into $z = (z_1, z_2)$, we can write

$$f(\tilde{p}|p^*)dF(p^*) = (T/2\pi)^{\frac{1}{2}nm} \exp(-\tfrac{1}{2}(z'z))dp^* = (2\pi)^{-\frac{1}{2}nm} T^{\frac{1}{2}n^*} \exp(-\tfrac{1}{2}(z'z))dz_2$$

where n^* is the number of elements in $\tilde{p}$, dz_2 is the differential element corresponding to z_2.

Substituting this in (30), we note that

$$z'z = \frac{1}{2}T\|p-\bar{p}\|^2 + \frac{1}{2}(z'z) \geq \frac{1}{2}T\delta_0^2 + \frac{1}{2}(z_2'z_2) , \quad \text{if} \quad \psi^*(p) \leq \epsilon^+ .$$

Thus from (30) we have

$$P(\psi^*(p)\leq\psi_0) \leq (2\pi)^{-\frac{1}{2}nm} \sum_L T^{\frac{1}{2}n^*} e^{-\frac{1}{2}T\delta_0^2}/B \int_{\omega_L:\psi(p)\leq\psi_0} d\sigma^* \exp -\frac{1}{4}(z_2'z_2)dz_2 .$$

Now for any $n^* < k^+$, we can find b_0, such that $T^{\frac{1}{2}n^*} e^{-\frac{1}{4}T\delta_0^2} < b_0$, if $T > T_0$ for some T_0. Thus

$$P(\psi^*(p) \leq \psi_0) \leq (2\pi)^{-\frac{1}{2}nm} b_0/B \sum_L \int_{\omega_L:\psi(p)\leq\psi_0} d\sigma^* \exp -\frac{1}{4}(z_2'z_2)dz_2 ,$$

and the preceding proof can now be repeated but the resulting bound is independent of T.

5. Two Examples of the Use of the Theorem

The first model considered is a simple example where the maximal finite moment is that indicated by Theorem 1. The model can be written

$$y_{1t} + \alpha_1 y_{2t} + \alpha_2 z_{1t} = u_{1t}$$

$$\alpha_2 y_{1t} + y_{2t} + \alpha_1 z_{2t} = u_{2t} .$$

In this case $n = m = N = 2$, so that if the zero set consists of only the full rank manifold we have $q = N-1 = 1$ and the maximal exponent should be less than or equal to 3. We can use Theorem 2 to show that maximal exponent is three. Assuming for simplicity that $M = I$, it is not difficult to obtain that

$$K' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

and that

$$K'(I \otimes R)K = \begin{pmatrix} 1 + p_3^2 + p_4^2 & p_3 + p_2 \\ p_3 + p_2 & 1 + p_1^2 + p_2^2 \end{pmatrix}$$

where
$$\hat{P} = \begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix}.$$

Thus $\psi^*(p) = p_1^2(1 + p_3^2 + p_4^2) + p_4^2(1 + p_2^2) + (1 - p_2p_3)^2$. It is clear that the zero set consists of points satisfying $p_1 = p_4 = 0$, $p_2p_3 = 1$, which is a one dimensional full-rank manifold. Taking $\xi_{10} = p_1$, $\lambda_1 = 1 + p_3^2 + p_4^2$, $\xi_{20} = p_4$, $\lambda_2 = 1 + p_2^2$, $\xi_{30} = p_2p_3 - 1$, $\lambda_3 = 1$, $r_i = 2$, $i = 1, 2, 3$, we can apply Theorem 2. Taking $\epsilon^+ = \frac{1}{4}$, Ω^+ can be divided up so that in Ω_1^* $\sqrt{2}|p_2| > 1$, and in Ω_2^* $\sqrt{2}|p_2| \leq 1$. In Ω_1^* we take $\tilde{p} = (p_1, p_3, p_4)$ and then $J = |p_2| \geq 1/\sqrt{2}$. In Ω_2^* we take $\tilde{p} = (p_1, p_2, p_4)$ and then $J = |p_3|$. Since $\psi^* \leq \frac{1}{4}$, $|\xi_{30}| \leq \frac{1}{2}$, so that $|p_3| \geq \frac{1}{2}/|p_2| \geq 1/\sqrt{2}$. Thus in any case $J \geq 1/\sqrt{2}$, and Theorem 2 shows that $E(\psi^*(p)^{-1/2\,k})$ is finite if $k < 3$. Thus from the previously discussed inequality the 3SLS estimates have maximal exponent 3.

The next application is considerably more complex and is designed to illustrate the special case of Section 3. The model can be written

$$(1+\alpha_1+\alpha_3)y_{1t} + (\alpha_2+\alpha_3)y_{2t} + \alpha_1 z_{1t} + (\alpha_2+\alpha_3)z_{2t} + (\alpha_3 - 3)z_{3t} = u_{1t}$$

$$(\alpha_1+\alpha_4)y_{1t} + (\alpha_2+\alpha_4-1)y_{2t} + \alpha_4 z_{1t} + (\alpha_2+\alpha_3)z_{2t} + (\alpha_4+2)z_{3t} = u_{2t} .$$

We then have

$$K' = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix} .$$

So that

$$K_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

and

$$K_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} ,$$

assuming that $M = I$. If we take

$$G_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} ,$$

we find that the conditions (14) and (15) are satisfied with

$$G_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad G_3 = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} , \quad G_4 = I_4 , \quad G_5 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} .$$

Here $m = 3$, $N_1 = 2$, $N_2 = 1$, so that this gives an upper bound on the maximal exponent of 2 . Now writing

$$\hat{P} = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \end{pmatrix}$$

we have that

$$E' = K'\left(I \otimes \begin{pmatrix} P \\ I \end{pmatrix} \right) = \begin{pmatrix} p_1+1 & p_2 & p_3 & p_1 & p_2 & p_3 \\ p_4 & p_5+1 & p_6 & p_4 & p_5+1 & p_6 \\ p_1+p_4 & p_2+p_5+1 & p_3+p_6+1 & 0 & 1 & 0 \\ 0 & 0 & 0 & p_1+p_4+1 & p_2+p_5 & p_3+p_6+1 \end{pmatrix}$$

Then $\det(E'E)$ can be somewhat simplified by carrying out elementary row operations on E' to obtain

$$(31) \quad \psi^*(p) = 4 \begin{vmatrix} p_1+\frac{1}{2} & p_2 & p_3+\frac{1}{2} \\ p_4 & p_5+1 & p_6 \\ -\frac{1}{2} & 1 & p_1+p_4+\frac{1}{2} \end{vmatrix}^2 + 4(\eta_1^2+\eta_2^2+\eta_3^2)((p_2+p_5)^2+(p_3+p_6+1)^2) ,$$

where $\eta_1 = p_2p_6 - (p_3 + \frac{1}{2})(p_5 + 1)$, $\eta_2 = p_4(p_3 + \frac{1}{2}) - p_6(p_1 + \frac{1}{2})$, $\eta_3 = (p_1 + \frac{1}{2})(p_5 + 1) - p_2p_4$. If any one of $p_1 + \frac{1}{2}$, p_2 , $p_3 + \frac{1}{2}$, p_4 , $p_5 + 1$, p_6 is non-zero, then it is possible to express this in terms of two of η_1 , η_2 , η_3 . The zero set consists of two manifolds of different dimensions. The first of these, already mentioned, is the intersection of $\eta_1 = \eta_2 = \eta_3 = 0$. This is of dimension 4 as can be seen from parameterization.

The second manifold of dimension 3 is the full rank manifold given by

$$p_2 + p_5 = 0$$

$$p_3 + p_6 \neq 1 = 0$$

$$\begin{vmatrix} p_1 + \frac{1}{2} & p_2 & p_3 + \frac{1}{2} \\ p_4 & p_5 + 1 & p_6 \\ -\frac{1}{2} & 1 & p_1 + p_4 + \frac{1}{2} \end{vmatrix} = 0 \ .$$

Using the first two equations, the last equation is easily shown to be equivalent to

$$\text{either} \quad p_1 + p_4 + 1 = 0$$

$$\text{or} \quad (p_1 + p_4)p_2 = p_1 + p_3 + 1 \ .$$

Thus the full rank manifold consists of two separate manifolds one of which is a linear subspace. In order to apply Theorem 2 it is necessary to consider sets Ω^+ which separately cover all three manifolds. A complete discussion would be very long, and in this section only Ω^+ covering the first manifold will be discussed, and some arguments will be very much summarized.

The manifold $\eta_1 = \eta_2 = \eta_3 = 0$ is a cone with vertex the point $(-1/2, 0, -1/2, 0, -1, 0)$. In order to discuss the integral of the p.d.f over a set Ω^+ covering this cone, it is necessary first to subtract a neighborhood of the vertex, and discuss separately the integral over this neighborhood. This can be done using a polar coordinate transformation. Omitting this section of the discussion consider now the integral over the set Ω^+ excluding the neighborhood. It now follows that we can partition Ω^+ into subsets such that on each subset one out of $(p_1 + 1/2, p_2, p_3 + 1/2, p_4, p_5 + 1, p_6)$ is non-zero throughout the subset. Consider for example the subset such that p_2 is non-zero throughout. Then take $\xi_{10} = \eta_1$, $\xi_{20} = \eta_3$, and note that $-p_2\eta_2 = (p_1 + 1/2)\xi_{10} + (p_3 + 1/2)\xi_{20}$.

However it is clear that if we define $b_1 = p_2 + p_5$, and $b_2 = p_3 + p_6 + 1$, we have the inequality

$$\psi^*(p) \geq 4(b_1^2 + b_2^2)(\eta_1^2 + \eta_3^2)$$

$$\geq 4(b_1^2\xi_{10}^2 + 4b_2^2\xi_{20}^2) .$$

We therefore consider taking $k_1 = k_2 = 1$, $\xi_{11} = b_1$, $\xi_{21} = b_2$, and the conditions of Theorem 2 are satisfied if $\partial\xi/\partial p$ is of rank 4 throughout the subset. Now

$$\frac{\partial\xi}{\partial p} = \begin{pmatrix} 0 & p_6 & -(p_5+1) & 0 & -(p_3+\frac{1}{2}) & p_2 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ p_5+1 & -p_4 & 0 & -p_2 & p_1+\frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

and by elementary operations it is possible to see that this is of rank 4 if $p_2 + p_5 + 1 \neq 0$ and $p_2 \neq 0$. We therefore split our subset into two sub-sets. In the first we take $|p_2 + p_5 + 1| > \epsilon_b$ throughout the set, and use the above set of ξ_{ij}. In the second subset having chosen $\epsilon_b < 1$, we can assume that $|b_1| > 1 - \epsilon_b$, and so use $\xi_{10} = \eta_1$, $\xi_{20} = \eta_3$, $k_1 = k_2 = 0$, and make use of $4(b_1^2 + b_2^2) > 4(1 - \epsilon_b)^2$.

Thus we need merely check that $\partial\xi/\partial p$ is of rank 2, which is satisfied if $p_2 \neq 0$.

Similar arguments apply to the subsets defined such that $p_3 + \frac{1}{2} \neq 0$, $p_5 + 1 \neq 0$, $p_6 \neq 0$. Now consider a subset defined so that p_2, $p_3 + \frac{1}{2}$, $p_5 + 1$, p_6 are all close to zero throughout the subset, but either $p_1 + \frac{1}{2}$ or p_4 is bounded away from zero. It will be found that we can take $\xi_{10} = \eta_2$, $\xi_{20} = \eta_3$, $k_1 = 0$, $k_2 = 0$, and the $\partial\xi/\partial p$ matrix will be of rank 2 throughout the subset.

This example illustrates the difficulty of using the Theorem 2 to discuss even a small model and makes it clear that a more general result of the type of Theorem 2 is badly needed to classify models into more general classes with respect to the existence of the moments of their 3SLS estimators.

6. A Lower Bound for the Separable Constraint Case

The previous discussion for the general linear constraint case is unsatisfactory in producing no general conclusions. However the situation is more satisfactory for the practically interesting case where each constraint refers to the coefficients of only one equation. Using similar inequalities to those of Section 5, we consider the coefficients of the j^{th} equation only by introducing an $n \times 1$ vector e_j, which has all its elements zero except the j^{th} element, which is one. Then starting from the inequality (27) of Section 4, and assuming as before that Ψ and Ψ^{-1} have all their elements bounded, we note $e_j e_j' - \lambda\Psi$ will be non-positive definite if $\lambda \geq e_j'\Psi^{-1}e_j$. Thus

$$\begin{aligned}(\hat{a}_j - \bar{a}_j)'R(\hat{a}_j - \bar{a}_j) &= \operatorname{tr}(e_j e_j'(\hat{A} - \bar{A})R(\hat{A} - \bar{A})') \\ &\leq (e_j'\Psi^{-1}e_j)\operatorname{tr}(\Psi(\hat{A} - \bar{A})R(\hat{A} - \bar{A})') \\ &\leq 4 \operatorname{tr}(\Psi\bar{A}R\bar{A}')(e_j'\Psi^{-1}e_j) .\end{aligned}$$

Now writing $a_j = a_{j0} + K_j^*\alpha_j$, we have

$$(\hat{\alpha}_j - \bar{\alpha}_j)'K_j^{*\prime}RK_j^*(\hat{\alpha}_j - \bar{\alpha}_j) \leq 4 \operatorname{tr}(\Psi\bar{A}R\bar{A}')e_j'\Psi^{-1}e_j) .$$

Defining λ_j^* as the smallest latent root of $K_j^{*\prime}RK_j^*$ we have

$$T(\hat{\alpha}_j' - \bar{\alpha}_j')(\hat{\alpha}_j - \bar{\alpha}_j) \leq 4T \operatorname{tr}(\Psi\bar{A}R\bar{A}')(e_j'\Psi^{-1}e_j)/\lambda_j^* .$$

Although as in the Theorem 2 it is possible to relax this assumption, for simplicity and in order to make use of the known theory of the non-central Wishart distribution we assume that all the u_t are independently distributed $N(0, \Omega_u)$. Using as before that $\lambda_j^* \geq \det(K_j^*RK_j^*)(\operatorname{tr}(K_j^{*\prime}RK_j^*)^{N_j-1}$ where $N_j = n + m - r_j$, and r_j is the number of constants on equation j, we have

$$(32) \quad T(\hat{\alpha}_j - \bar{\alpha}_j)'(\hat{\alpha}_j - \bar{\alpha}_j) \leq 4T \operatorname{tr}(\Psi\bar{A}R\bar{A})(e_j'\Phi^{-1}e_j)(\operatorname{tr}\Phi_j)^{N_j-1}/\det \Phi_j$$

where $\Phi_j = K_j^{*'} R K_j^*$. Define

$$\zeta_j(p) = 4T \; tr(\Psi \bar{A} R \bar{A}')(e_j' \Psi^{-1} e_j)(tr(\Phi_j))^{N_j - 1}$$

so that

$$E([T(\hat{\alpha}_j - \bar{\alpha}_j)'(\hat{\alpha}_j - \bar{\alpha}_j)]^{\frac{1}{2}k}) \le E((\zeta_j(p))^{\frac{1}{2}k} / \det^{\frac{1}{2}k}(\Phi_j)) \; .$$

Since we have assumed that the largest and smallest latent roots of Ψ are bounded, it is simplest in the rest of this section to treat Ψ as constant, so that the dependence of the expression above on w can be neglected, so that the last expectation can be written

$$(T/2\pi)^{\frac{1}{2}nm} (\det \Omega_v)^{-\frac{1}{2}m} \int (\zeta_j(p))^{\frac{1}{2}k} (\det \Phi_j)^{-\frac{1}{2}k} \exp(-\tfrac{1}{2}T\,(p-\bar{p})'(\Omega_v^{-1} \otimes I)(p-\bar{p})dp \; .$$

Consider now the function

$$\zeta_j(p)\exp(-T/2k(\bar{p}-\bar{p})'(\Omega_v^{-1} \otimes I)(p-\bar{p})) \; . \qquad (33)$$

This is continuous for all finite p , and tends to zero at infinity, and so has an upper bound in p space. Noting that $T(\bar{A}R\bar{A}') = T\bar{B}(\hat{P}-\bar{P})(\hat{P}-\bar{P})'\bar{B}'$, where $\bar{P} = E(\hat{P})$, which remains $O(1)$ as $T \to \infty$, it is not difficult to see that as $T \to \infty$, the function (33) is dominated by

$$k^* \; tr(\Psi \bar{A} R \bar{A}')\exp(-T/2k(\bar{p}-\bar{p})'(\Omega_v^{-1} \otimes I)(p-\bar{p})) \; ,$$

for some large but finite k^* , so that the bound on the function remains finite as $T \to \infty$. Let the uniform bound be denoted by M_{jk} . Then

$$E((T(\hat{\alpha}_j - \bar{\alpha}_j)'(\hat{\alpha}_j - \bar{\alpha}_j))^{\frac{1}{2}k})$$
$$\leq M_{jk}^{\frac{1}{2}k}(T/2\pi)^{\frac{1}{2}nm}(\det \Omega_v)^{-\frac{1}{2}m}\int(\det \Phi_j)^{-\frac{1}{2}k}\exp(-\tfrac{1}{4}T(p-\bar{p})'(\Omega_v^{-1} \times I)(p-\bar{p}))dp$$
$$= 2^{\frac{1}{2}nm}E^*(\det(\Phi_j)^{-\frac{1}{2}k})M_{jk}^{\frac{1}{2}k},$$

where E^* denotes the expectation taking the variance matrix of $(p-\bar{p})$ to be that obtained when the variance matrix Ω_v is double its original value. The advantage of this inequality is that Φ_j is a non-central Wishart matrix, so that known results on the moments of the determinant of such a matrix can be used. We can write $\Phi_j = D_j^* D_j^{*'}$, where

$$D_j^* = K_{j1}'(Y'Z)(Z'Z)^{-1/2}/\sqrt{T} + K_{j2}'$$
$$= K_{j1}'(V'Z)(Z'Z)^{-1/2}/\sqrt{T} + (K_{j1}'\bar{P} + K_{j2}').$$

Thus $E(D_j^*) = K_{j1}'\bar{P} + K_{j2}'$, and its columns are independently and normally distributed with variance matrix $(K_{j1}'\Omega_v K_{j1})/T$. It is impossible to use the results of Herz [1] directly if this variance matrix is singular. Suppose then that its rank is $N_j^+ \leq N_j = m+n-r_j$. We can find an orthogonal matrix H_j^+ such that

$$H_j^+(K_{j1}'\Omega_v K_{j1})H_j^{+'} = \begin{pmatrix} \Omega_j^* & 0 \\ 0 & 0 \end{pmatrix},$$

and Ω_j^* is an $N_j^+ \otimes N_j^+$ positive definite matrix. Define $D_j^{+'} = (H_j^+ D_j^*)'$ $= (D_{j1}^{+'} : D_{j2}^{+'})$, and note that D_{j2}^+ is non-stochastic. Define an $m \times (m - N_j + N_j^+)$ matrix H_j^*, so that its columns are mutually orthogonal and so that $H_j^* H_j^{*'} = I - D_{j2}^{+'}(D_{j2}^{+}D_{j2}^{+'})^{-1}D_{j2}^{+}$, and then define $\tilde{D}_j = D_{j1}^{+}H_j^*$. Then $\det(D_j^* D_j^{*'}) = \det(D_j^{+}D_j^{+'}) = \det(D_{j2}^{+}D_{j2}^{+})\det(\tilde{D}_j\tilde{D}_j')$. $\det(D_{j2}^{+}D_{j2}^{+})$ is a constant (non-stochastic and independent of T), which is strictly positive if the equation j is identified. Then $E(\tilde{D}_j) = H_{j1}^{+}(K_{j1}'P + K_{j2}')H_j^*$, where H_{j1}^{+} consists of the first N_j^{+} rows of H_j^{+}. Define

$$\bar{\Phi}_j = E(\tilde{D}_j)E(\tilde{D}_j') = H_{1j}^{+}(K_{j1}'\bar{P} + K_{j2}')(I - D_{j2}^{+'}(D_{j2}^{+}D_{j2}^{+})^{-1}D_{j2}^{+})(\bar{P}'K_{j1}^{+} + K_{j2}^{+})H_{1j}^{+'},$$

and note that the columns of $\tilde{D}_j$ are independently identically distributed normally with variance matrix Ω_j^*/T. We can then use the results of Herz [1], that

$$E((\det \tilde{D}_j\tilde{D}_j')^{-\frac{1}{2}k})$$

$$= \Gamma_{N_j^+}(\tfrac{1}{2}(m - N_j + N_j^+ - k))/\Gamma_{N_j^+}(\tfrac{1}{2}(m - N_j + N_j^+))[\det(2\Omega_j^*/T)]^{-\frac{1}{2}k}\, {}_1F_1(\tfrac{1}{2}k, \tfrac{1}{2}(m - N_j + N_j^+); -\tfrac{1}{2}T\bar{\Phi}_j\Omega_j^{*-1})$$

This result was only discussed by Herz for the case where k is negative, but it is easy to see that his proof extends to the case where

$$(34) \qquad k < m - N_j + 1 .$$

From the definition of the hypergeometric function

$$E((\det(\tilde{D}_j\tilde{D}_j'))^{-\frac{1}{2}k} = \frac{(\det \bar{\Phi}_j)^{-\frac{1}{2}k}}{\Gamma_{N_j^+-1}(\frac{1}{2}k)} \int_0^{TS_m} \operatorname{etr}(-S)(\det S)^{\frac{1}{2}(k-N_j^+)} \det(I - SS_m^{-1}/T)^{\frac{1}{2}(m-k-N_j^+)} dS$$

where $S_m = \frac{1}{2}(\bar{\Phi}_j^{1/2}\Omega_j^{*-1}\bar{\Phi}_j^{1/2})$, and the range of integration of S is all non-negative definite matrices such that $TS_m - S$ is non-negative definite. Note that if S_m is positive definite, we can find a sufficiently large neighborhood of the origin, such that, as $T \to \infty$, the contribution of the integral outside the neighborhood can be neglected, and that within the neighborhood the third term in the integrand can be put equal to one, so that

$$\lim_{T\to\infty} E(\det(\tilde{D}_j\tilde{D}_j'))^{-\frac{1}{2}k} = (\det \bar{\Phi}_j)^{-\frac{1}{2}k} ,$$

and the bound on the moment is uniform as $T \to \infty$.

7. <u>The Case Where 2SLS is Used to Estimate</u> Ω_u

This last section develops an upper bound for the case where Ω_u is estimated by 2SLS. I conjecture that the correct maximal exponent for this case is the same as for the case Ψ constant, but I have not been able to prove this.

Considering only the separable constraint case, denote the 2SLS estimates of the matrix A by $\tilde{A}$, suppose that we are interested in the 3SLS estimates of the first equation, so that we write $A' = (a_1, A_2')$, and define $A^{*\prime} = (\hat{a}_1, \tilde{A}_2')$ where as before $\hat{}$ means 3SLS estimates, and $\sim$ means 2SLS estimates. Define

$$\tilde{\Omega} = \tilde{A}(X'X)\tilde{A}'/T = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix} .$$

From the minimization definition of the 3SLS estimates we have

$$\mathrm{tr}(\tilde{\Omega}^{-1}\hat{A}R\hat{A}') \leq \mathrm{tr}(\tilde{\Omega}^{-1}A^{*}RA^{*\prime}) .$$

Then as in the last section we can deduce that

$$(\hat{a}_1 - \bar{a}_1)'R(\hat{a}_1 - \bar{a}_1) \leq 4\omega_{11}\ \mathrm{tr}(\Omega^{-1}A^*RA^{*\prime}) \ .$$

Now write

$$\Omega^{-1} = \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \Psi_{22} \end{pmatrix} ,$$

so that

$$\begin{aligned} \mathrm{tr}(\Omega^{-1}A^*RA^{*\prime}) &= \psi_{11}(\bar{a}_1'R\bar{a}_1) + 2(\psi_{12}\tilde{A}_2R\bar{a}_1) + \mathrm{tr}(\Psi_{22}\tilde{A}R\tilde{A}') \\ &\leq [(\psi_{11}(\bar{a}_1'R\bar{a}_1))^{\frac{1}{2}} + [(\psi_{12}\tilde{A}_2R\tilde{A}_2'\psi_{21})/\psi_{11}]^{\frac{1}{2}}]^2 + \mathrm{tr}(\Omega_{22}^{-1}\tilde{A}_2R\tilde{A}_2') \ , \end{aligned}$$

where use is made of the Cauchy inequality,

$$\sqrt{(\bar{a}_1'R\bar{a}_1)(\psi_{12}\tilde{A}R\tilde{A}'\psi_{21})} \geq (\bar{a}_1'R\tilde{A}'\psi_{21}) \ ,$$

and that $\Psi_{22} = \psi_{21}\psi_{12}/\psi_{11} + \Omega_{22}^{-1}$. Also we have

$$\begin{aligned} \tilde{A}_2R\tilde{A}_2' &= \tilde{A}_2(X'X)\tilde{A}_2'/T - \tilde{A}_2(X'(I - Z(Z'Z) \quad Z')X)\tilde{A}_2'/T \\ &\leq \Omega_{22} \ , \end{aligned}$$

so that, $\mathrm{tr}(\Omega_{22}^{-1}A\ RA') \leq n-1$, and $\psi_{12}\tilde{A}_2'R\tilde{A}_2'\psi_{21} \leq \psi_{12}\Omega_{22}\psi_{21} = -\psi_{12}\omega_{21}\psi_{11}$ $= \psi_{11}(1 - \psi_{11}\omega_{11})$. Thus

$$(\Omega^{-1}A^*RA^{*\prime}) < [(\psi_{11}(\bar{a}_1'R\bar{a}_1))^{\frac{1}{2}} + \sqrt{1 - \psi_{11}\omega_{11}}]^2 + n - 1$$

and so

$$(\hat{a}_1 - \bar{a}_1)'R(\hat{a}_1 - \bar{a}_1) \le 4[((\psi_{11}\omega_{11})^{\frac{1}{2}}(\bar{a}_1'R\bar{a}_1)^{\frac{1}{2}} + (\omega_{11}(1 - \psi_{11}\omega_{11}))^{\frac{1}{2}})^2 + (n-1)\omega_{11}] .$$

Also $0 \le \psi_{11}\omega_{11} \le 1$, so that finally

$$(\hat{a}_1 - \bar{a}_1)'R(\hat{a}_1 - \bar{a}_1) \le 4((\bar{a}_1'R\bar{a}_1) + 2\sqrt{\omega_{11}(\bar{a}_1'R\bar{a}_1)} + n\omega_{11} .$$

Now

$$\omega_{11} = u_1'u_1/T + 2(u_1'X)(\tilde{a}_1 - \bar{a}_1)/T + (\tilde{a}_1 - \bar{a}_1)'(X'X)(\tilde{a}_1 - \bar{a}_1)/T ,$$

so that using the Cauchy inequality again we have

$$\sqrt{\omega_{11}} \le \sqrt{u_1'u/T} + ((\tilde{a}_1' - \bar{a}_1')(X'X)(\tilde{a}_1 - \bar{a}_1)/T)^{1/2} .$$

Writing $w_1 = (Z'Z)^{-1/2}(Z'u_1)/\sqrt{T}$, and $H_1 = (K_1^{*\prime}RK_1^*)^{-1}K_1^{*\prime}(X'Z)(Z'Z)^{-1/2}/\sqrt{T}$, we have that $\tilde{a}_1 - \bar{a}_1 = K_1^*(\tilde{\alpha}_1 - \bar{\alpha}_1) = K_1^*H_1w_1$.

Then

$$\begin{aligned}(\tilde{a}_1 - \bar{a}_1)'(X'X)(\tilde{a}_1 - \bar{a}_1)/T &= tr(H_1'[(K_1^{*\prime}X'XK_1^*)/T]H_1w_1w_1') \\ &\le tr(H_1'(K_1^{*\prime}X'XK_1^*/T)H_1)(w_1'w_1) ,\end{aligned}$$

since $w_1'w_1$ is the largest latent root of the matrix w_1w_1'

$$\begin{aligned}&= tr((K_1^{*\prime}RK_1^*)^{-1}K_1^{*\prime}X'XK_1^*/T)(w_1'w_1) \\ &\le \lambda_1^{*-1}\ tr(K_1^{*\prime}X'XK_1^*/T)(w_1'w_1) ,\end{aligned}$$

where λ_1^* is the smallest latent root of $K_1^{*\prime}RK_1^*$. Writing $d_1^2 = tr(K_1^{*\prime}X'XK_1^*/T(w_1'w_1)$ and $s_1 = \sqrt{u_1'u_1/T}$, we have

$$(\hat{a}_1 - \bar{a}_1)(\hat{a}_1 - \bar{a}_1) \leq 4/\lambda_1^*[\bar{a}_1'R\bar{a}_1 + 2(\bar{a}_1'R\bar{a}_1)^{\frac{1}{2}}(s_1 + d_1\lambda_1^{*-\frac{1}{2}}) + n(s_1 + d_1\lambda_1^{*-\frac{1}{2}})^2]$$

$$= 4/\lambda_1^{*2}[\lambda_1^*(\bar{a}_1'R\bar{a}_1) + 2(\lambda_1^*(\bar{a}_1'R\bar{a}_1))^{\frac{1}{2}}(s_1\lambda_1^{*\frac{1}{2}} + d_1) + n(s_1\lambda_1^{*\frac{1}{2}} + d_1)^2] \ .$$

At this stage the argument can be continued as in the last section, since the factor in the square brackets can be dominated in the integration in a similar way. The significant difference is that λ_1^{*2} now occurs in the denominator, so that the previous argument shows that the k^{th} moments of $(\hat{a}_1 - \bar{a}_1)$ are finite if

(35) $$2k < m - N_j + 1 \ .$$

This is, of course, only a sufficient condition for the existence of the moment, and is probably not necessary. However for low values of k, it provides a useful result, showing that the mean exists if the degree of overidentification is two, and that the variance exists if the degree of overidentification is four. Unfortunately the proof does not show that the bounds are uniform as $T \to \infty$, so that the result cannot be used to validate a Nagar expansion as in my article [3].

8. General Comments and Conclusions

The results of this article are incomplete in not giving definite answers either in the general case, or in the separable case with 2SLS variance matrix. The bounds obtained are generally crude but are sufficient in most cases to allow Nagar approximations to be developed [3], which are the most useful method for analyzing the factors affecting the lower order moments of these estimators. Unfortunately the last section does not give the uniform bounds as $T \to \infty$ required for this purpose.

BIBLIOGRAPHY

[1] Herz, C. S. (1955) "Bessel Functions of Matrix Argument," Ann. of Math., Vol. 61, pp. 474-523.

[2] Hu, S. T. (1969) Differential Manifolds. New York: McGraw Hill.

[3] Mariano, R. S. (1972) "The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators," Econometrica, Vol. 40, pp. 643-652.

[4] Sargan, J. D. (1974) "On the Validity of Nagar's Expansion for the Moments of Econometric Estimators," Econometrica (forthcoming).

[5] Zellner, A. and H. Theil (1962) "Three-Stage Least-Squares: Simultaneous Estimation of Simultaneous Equations," Econometrica, Vol. 30, pp. 54-78.