

Noncentral Student distributed LS and IV Estimators.

Leon L. Wegge¹

The distribution of the least squares and the instrumental variable estimators of the coefficients in a linear relation is noncentral student when the data are normally distributed around possibly non-constant means. This is the claim of the paper and we show for what definition of a noncentral student density this claim to compact summary of the vast literature is justified. Unfortunately, the definition of the noncentral student density is as complicated as its parent, the noncentral Wishart density. Both are defined in terms of infinite series of zonal polynomials of all orders. We have developed however a recursive online method that generates these polynomials sequentially ad infinitum for bivariate and trivariate densities. The time is here that the practicality of the theory can be widened considerably.

0. Introduction.

The practitioner of statistical and econometric methods often deals with data that are truly random in the sense that the researcher has no control over their generation, there being no design of any kind. Yet, to evaluate the evidence, all we have are either theorems that pertain to designed data or are large sample probability statements. The latter are expressions of what would be if we had a sample of infinite size. There is no basis to their relevance.

An assessment of the relevance of the large sample properties of estimators in finite samples is possible if their finite sampling distribution is known. This paper is a summary of the literature on the distribution of the least squares and instrumental variable estimators of linear equation coefficients and sketches a view towards unification and practical implementation of the results.

This work started with James [3], [4], with his definition of the noncentral Wishart density and his derivation that the moment matrix of normally distributed data with shifting means is noncentral Wishart distributed. Polynomials of all orders, called zonal polynomials, appear in that definition and these appear to be difficult to compute. Parkhurst & James [8] calculated zonal polynomials up to order 12. For practical purposes polynomials of much higher order are needed, since the polynomials converge very slowly.

An important contribution to the theory of the noncentral Wishart density was made by Constantine [1] by showing the reproductive property of zonal polynomials, i.e. the expectation of a zonal polynomial under a central Wishart distribution is another zonal polynomial.

Student densities, both central and noncentral, have a much longer history. There are many definitions. Traditionally these densities were constructed as multivariate generalizations of standard univariate student t-ratios. In the context of the literature around the Wishart density there is a natural way that leads to a multivariate student density concept. With normal data and the moment matrix following a central Wishart density, what is the density of the least squares estimator of one subset of the data on another subset?

Kshirsagar [7] asked this question and he showed that the least squares estimator follows a multivariate central student density. Extensions to include some fixed data in the regression model were published in [6] and [13].

In this paper we ask the same question, but in the context of the noncentral Wishart density. When the moment matrix is noncentral Wishart distributed, what is the density of the least squares estimator? At the end of the analysis, we can define the noncentral student density in such a way that the noncentral Wishart and student densities are linked in the same way as was discovered by Kshirsagar in the case of the central Wishart and student densities. This link is recoverable through the derivation of the density of the least squares estimator in either case.

¹ Professor of Economics emeritus, University of California, Davis.

As is well known the densities of the least squares estimator and the instrumental variable estimator have many features in common [9]. It is not surprising to find that the noncentral student density also characterizes the density of the instrumental variable estimator. The difference in the degrees of freedom and the inequality between the coefficient in the linear relation and the population regression coefficient create the difference in the density parameters.

When the transformations and the algebraic derivations between the noncentral Wishart and noncentral student densities are done, the question of the implementation of the results remains. Phillips [9] concluded with the summary statement that the joint densities satisfy a functional differential equation. Here the joint densities are characterized as noncentral student densities that depend on analytically intractable hypergeometric functions that are series of zonal polynomials of all orders.

There is progress however in the online calculation of the zonal polynomials that appear in bivariate and trivariate densities. We provide the details in an appendix. We stand on the view that the results are not only satisfying from the theory perspective, but also practical in these lower dimensional distributions.

The paper is divided in 4 sections. Section 1 defines the noncentral Wishart density and states the fact that the moment matrix has this distribution if the data come from a normal population with changing means. In Section 2 we define the central student density, the moment matrix is partitioned and the transformation from the moment matrix to i) the least squares estimators (LSE), to ii) the moment matrix of regressors and iii) to the residual variance is performed. In the noncentral situation, the latter two have interdependent Wishart densities and this creates a main problem. The density of the LSE is found to be a product of a central student density and is also linked to the Wishart densities through the hypergeometric function. This function links all three statistics jointly. Section 3 studies the marginal joint density of the LSE and interprets the noncentral student density as a weighted average of central student densities. We show that the LSE of the coefficients in a linear stochastic relation between stochastic regressand and regressors that have nonstationary means is distributed in a multivariate noncentral student distribution. In Section 4 we show the comparison between the instrumental variable estimator (IVE) and the LSE. The IVE is found to have a noncentral student density. This result also pertains to what is known in the econometric literature as indirect least squares estimators. Section 5 concludes by raising the potential for practical uses of the theory because progress in the computational methods has been made. The Appendix provides details on a recursive method for calculating zonal polynomials.

1. The Non-central Wishart Density.

Suppose that the columns of the $p \times n$ matrix X are independently normally distributed with covariance matrix Σ and expected value $E(X) = M$. The probability density is

$$(2\pi)^{-pn/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \text{tr}(\Sigma^{-1}(X - M)(X - M)')\right]. \quad (1)$$

The density of the $p \times p$ matrix $S = XX'$ is given by James [4] as

$$f_w(S|p, \Sigma, n, M) = w(p, \Sigma, n, M) |S|^{(n-p-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\Sigma^{-1}S)\right] {}_0F_1\left(\frac{n}{2}; \frac{1}{4} \Sigma^{-1}MM' \Sigma^{-1}S\right). \quad (2)$$

where

$$w(p, \Sigma, n, M) = (\Gamma_p(\frac{n}{2}))^{-1} |2\Sigma|^{-n/2} \exp[-\frac{1}{2} \text{tr}(\Sigma^{-1} M M')] , \quad (3)$$

$\Gamma_p(\frac{n}{2}) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma[\frac{1}{2}(n+1-i)]$ and ${}_0F_1(\frac{n}{2}; D)$ is the hypergeometric function. If $Z_\kappa(D)$ is a up to length p James zonal polynomial of order k , Constantine [1] renormalized the polynomials. He defined $C_\kappa(D) = [\chi_{[2\kappa]}(1) 2^k k! / (2k)!] Z_\kappa(D)$ and then

$${}_0F_1(\frac{n}{2}; D) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\kappa} \frac{C_\kappa(D)}{(n/2)_\kappa}, \quad \text{with partitions } \kappa = (k_1, k_2, \dots, k_p), \sum_{i=1}^p k_i = k, (n/2)_\kappa = \prod_{i=1}^p (\frac{n}{2} - \frac{1}{2}(i-1))_{k_i}$$

$$(x)_{k_i} = x(x+1)\dots(x+k_i-1) \text{ and } \chi_{[1\kappa]}(1) = k! \prod_{i < j} (k_i - k_j - i + j) / \prod_{i=1}^p (k_i + p - i)!$$

The density (2) is called the non-central Wishart density on n degrees of freedom with matrix of non-centrality parameters $\frac{1}{2} M M' \Sigma^{-1}$. Wishart [14] found (2) when $M = 0$.

2. The multivariate student density .

Folowing an outline introduced in [13], partition the $p \times p$ matrices

$$S = \begin{pmatrix} S_{11} & S_{11}b \\ b' S_{11} & S_{22.1} + b' S_{11}b \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} (\Sigma_{11})^{-1} + \beta(\Sigma_{22.1})^{-1}\beta' & -\beta(\Sigma_{22.1})^{-1} \\ -(\Sigma_{22.1})^{-1}\beta' & (\Sigma_{22.1})^{-1} \end{pmatrix},$$

where S_{11} and Σ_{11} are $p_1 \times p_1$ principal submatrices, $\beta = (\Sigma_{11})^{-1} \Sigma_{12}$, $b = (S_{11})^{-1} S_{12}$, $S_{22.1} = S_{22} - S_{12}' (S_{11})^{-1} S_{12}$ and $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{12}' (\Sigma_{11})^{-1} \Sigma_{12}$. With $p_2 = p - p_1$, the $p_1 \times p_2$ matrix b is the least squares regression coefficient matrix, regressing the bottom p_2 variables on the top p_1 variables. We seek the density of b . In terms of the partitioned matrices and $M' = (M_1' M_2')$, M_1 having p_1 rows, observe that

$$\text{tr} \Sigma^{-1} S = \text{tr}[\Theta S_{11}] + \text{tr}[(\Sigma_{22.1})^{-1} S_{22.1}], \quad \text{with } \Theta = (\Sigma_{11})^{-1} + (b - \beta)(\Sigma_{22.1})^{-1}(b - \beta)', \quad (4)$$

$$\text{tr} M' \Sigma^{-1} M = \text{tr} M_1' (\Sigma_{11})^{-1} M_1 + \text{tr} (M_2 - \beta M_1)' (\Sigma_{22.1})^{-1} (M_2 - \beta M_1) \quad (5)$$

$$M' \Sigma^{-1} S \Sigma^{-1} M = \Omega_1' S_{11} \Omega_1 + \Omega_2' S_{22.1} \Omega_2, \quad \text{with} \quad (6)$$

$$\Omega_1 = (\Sigma_{11})^{-1} M_1 + (b - \beta)(\Sigma_{22.1})^{-1} (M_2 - \beta M_1), \quad \Omega_2 = (\Sigma_{22.1})^{-1} (M_2 - \beta M_1).$$

Consider the variables in the matrix $S^0 = \begin{pmatrix} S_{11} & b \\ b' & S_{22.1} \end{pmatrix}$. The Jacobian of the transformation of the elements on and above the diagonal of S to S^0 is equal to $|S_{11}|^{p_2}$. From the density (2) of S and the equations (4) and (5) we find that the density of S^0 consists of three factors

$$f(S^0) = f_w(S_{11}|p_1, \Theta^{-1}, n+p_2, 0) \times f_w(S_{22.1}|p_2, \Sigma_{22.1}, n-p_1, 0) \quad (7)$$

$$\times \frac{w(p, \Sigma, n, M)}{w(p_1, \Theta^{-1}, n+p_2, 0) w(p_2, \Sigma_{22.1}, n-p_1, 0)} {}_0F_1\left(\frac{n}{2}; \frac{1}{4}(\Omega_1' S_{11} \Omega_1 + \Omega_2' S_{22.1} \Omega_2)\right),$$

with

$$f_w(S_{11}|p_1, \Theta^{-1}, n+p_2, 0) = w(p_1, \Theta^{-1}, n+p_2, 0) |S_{11}|^{(n+p_2-p_1-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\Theta S_{11})\right], \quad (8a)$$

$$f_w(S_{22.1}|p_2, \Sigma_{22.1}, n-p_1, 0) = w(p_2, \Sigma_{22.1}, n-p_1, 0) |S_{22.1}|^{(n-p-1)/2} \exp\left\{-\frac{1}{2} \text{tr}[(\Sigma_{22.1})^{-1} S_{22.1}]\right\}. \quad (8b)$$

The first two factors in (7) are central Wishart densities. The third factor is seen to equal

$$f_S(b|p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n+p_2, 0) \times \exp\left\{-\frac{1}{2} \text{tr}[M_1' (\Sigma_{11})^{-1} M_1]\right\}$$

$$\times \exp\left\{-\frac{1}{2} \text{tr}[(M_2 - \beta M_1)' (\Sigma_{22.1})^{-1} (M_2 - \beta M_1)]\right\} \times {}_0F_1\left(\frac{n}{2}; \frac{1}{4}(\Omega_1' S_{11} \Omega_1 + \Omega_2' S_{22.1} \Omega_2)\right), \quad (9)$$

where

$$f_S(b|p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n+p_2, 0)$$

$$= \frac{\Gamma_{p_2}\left(\frac{n+p_2}{2}\right)}{(\pi)^{p_1 p_2 / 2} \Gamma_{p_2}\left(\frac{n+p_2-p_1}{2}\right)} \frac{|S_{11}|^{-n/2} |S_{22.1}|^{-p_1/2}}{\left|(\Sigma_{11})^{-1} + (b - \beta)(\Sigma_{22.1})^{-1}(b - \beta)\right|^{(n+p_2)/2}} \quad (10a)$$

$$= \frac{|S_{11}|^{p_2/2} |S_{22.1}|^{(n+p_2-p_1)/2} \prod_{i=1}^{p_2} \Gamma\left[\frac{1}{2}(n+p_2+1-i)\right]}{(\pi)^{p_1 p_2 / 2} \prod_{i=1}^{p_2} \Gamma\left[\frac{1}{2}(n+p_2-p_1+1-i)\right] |S_{22.1} + (b - \beta)' \Sigma_{11} (b - \beta)|^{(n+p_2)/2}} \quad (10b)$$

is the multivariate central student density with parameters $(p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n+p_2, 0)$. q -th order moments exist provided $q \leq n - p_1$ and then

$$E(b) = \beta, \quad E(dd') = \frac{1}{n - p_1 - 1} (\Sigma_{22.1} \otimes (\Sigma_{11})^{-1}), \quad \text{where } d = \text{vec}(b - \beta). \quad (11)$$

The second equality (10b) follows from (10a) and the determinantal identity $|I_{p_1} + FG| = |I_{p_2} + GF|$, for $p_1 \times p_2$ F and G .

The complete noncentral density $f_S(b|p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n+p_2, M)$ of the least squares estimator b depends on the first two factors and is obtained from (7) by integrating out S_{11} and $S_{22.1}$. This density is a product of densities in which the central student density (10) is always a factor.

Example 1. Stationary Normal Regressors.

When $M=0$, the density of b is given by (10) since the hypergeometric function ${}_0F_1(\ ;1) = 1$ and the first two probability factors in (7) are independent of b integrating to one. This result was obtained by Kshirsagar [7]. The first two moments are stated above. In particular b is an unbiased estimator of the population regression coefficients $\beta = (\Sigma_{11})^{-1}\Sigma_{12}$. In the $p_2 \times n$ regression model $X_2 = \gamma' X_1 + U$,

where U are unobserved (residual) random variables and the columns of $\begin{pmatrix} X_1 \\ U \end{pmatrix}$ are independently normally distributed with means zero and covariance $\begin{pmatrix} \Sigma_{11} & \Sigma_{1v} \\ \Sigma_{v1} & \Sigma_{vv} \end{pmatrix}$, the data covariance

$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11}\gamma + \Sigma_{1v} \\ \Sigma_{12}' & \Sigma_{vv.1} + \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12} \end{pmatrix}$, the population regression coefficients $\beta = \gamma + (\Sigma_{11})^{-1}\Sigma_{1v}$ and b is a biased estimator of γ except when $\Sigma_{v1} = 0$, i.e. the residual variables U and the regressor variables X_1 are independent.

The result extends to the model $X_2 = \gamma' X_1 + \delta' Z + U$, where $p_3 \times n$ Z are fixed variates [6], [13] and the means of X_1 are linear forms in the fixed variates Z .

3. LSE b is multivariate non-central Student . $M_2 = \gamma' M_1, \gamma = \beta$.

In the general case with nonstationary means M , the $p_2 \times n$ regression model $X_2 = \gamma' X_1 + U$, where the columns of U are identically distributed independent normal unobserved residuals with mean zero, constant variance and distributed also independently of the regressors X_1 , implies the relation $M_2 = \gamma' M_1$ between means and $\Sigma_{21} = \gamma' \Sigma_{11}$ between the covariances, since the expectation of the second moments $E(X_2 X_1')$ must equal $\gamma' E(X_1 X_1') = \gamma' (\Sigma_{11} + M_1 M_1')$. The model implies $\beta = (\Sigma_{11})^{-1}\Sigma_{12} = \gamma$. Good regressors imply that property.

Substituting these equalities into (7), the variates $S_{22.1}$ are not present in the hypergeometric function. After integrating out $S_{22.1}$, the joint density of b and S_{11} is given by

$$f(b, S_{11}) = f_w(S_{11} | p_1, \Theta^{-1}, n + p_2, 0) \times {}_0F_1\left(\frac{n}{2}; \frac{1}{4} (M_1' (\Sigma_{11})^{-1} S_{11} (\Sigma_{11})^{-1} M_1)\right) \times f_s(b | p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n + p_2) \times \exp\left\{-\frac{1}{2} \text{tr}[M_1' (\Sigma_{11})^{-1} M_1]\right\} \quad (11)$$

Constantine [1] showed that the expectation in the central Wishart distribution of the hypergeometric function is a hypergeometric function i.e.

$$\int_{S_{11} > 0} \exp\left[-\frac{1}{2} \text{tr}(\Theta S_{11})\right] |S_{11}|^{(n+p_2-p_1-1)/2} {}_0F_1\left(\frac{n}{2}; \frac{1}{4} (\Sigma_{11})^{-1} M_1 M_1' (\Sigma_{11})^{-1} S_{11}\right) dS_{11} = \frac{2^{p_1(n+p_2)/2} \Gamma_{p_1}\left(\frac{n+p_2}{2}\right)}{|\Theta|^{(n+p_2)/2}} {}_1F_1\left(\frac{n+p_2}{2}; \frac{n}{2}; \frac{1}{2} (\Sigma_{11})^{-1} M_1 M_1' (\Sigma_{11})^{-1} \Theta^{-1}\right), \quad (12)$$

where ${}_1F_1(a; b; D) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a)_{\kappa} C_{\kappa} D}{(b)_{\kappa} k!}$, $\kappa = (k_1, k_2, \dots, k_{p_1})$.

Substituting into (11), the density of the least squares estimator b is

$$f_S(b|p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n + p_2, M_2 = \beta M_1) = \exp\left[-\frac{1}{2} \text{tr} M_1 M_1' (\Sigma_{11})^{-1}\right] \times {}_1F_1\left(\frac{n + p_2}{2}; \frac{n}{2}; \frac{1}{2} (\Sigma_{11})^{-1} M_1 M_1' (\Sigma_{11})^{-1} \Theta^{-1}\right) \times f_S(b|p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n + p_2, 0) \quad (13)$$

We will say that b is distributed in the noncentral student density with parameter set $(p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n + p_2, M_2 = \beta M_1)$.

Univariate Case.

The name noncentral student density is simply reflecting the fact that in the scalar simple regression case with $p_1 = p_2 = 1$, (13) is a weighted average of central student densities in the form

$$f_S(b|1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n + 1, M_2 = \beta M_1) = \sum_{k=0}^{\infty} \lambda_k f_S(b|1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n + 1 + 2k, 0), \quad \text{with } \lambda_k = \frac{1}{k!} \left(\frac{M_1 M_1'}{2 \Sigma_{11}}\right)^k / \exp\left[\frac{M_1 M_1'}{2 \Sigma_{11}}\right]. \quad (14)$$

The mean $E(b) = \beta$ and $\text{var } b = \sum_{k=0}^{\infty} \lambda_k \frac{\Sigma_{22.1}}{n + 2k - 2} (\Sigma_{11})^{-1}$, which is smaller than that of the central distribution.

Multivariate Case.

$$f_S(b|1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n + 1, M_2 = \beta M_1) = \sum_{k=0}^{\infty} \mu_k(b) f_S(b|1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n + 1 + 2k, 0), \quad \text{with } \mu_k(b) \text{ dependent on } b. \quad (15)$$

Proof.

i). We prove (15) first for the single equation multiple regression case when $p_2 = 1$, p_1 is arbitrary and $M_1 M_1' (\Sigma_{11})^{-1} = I_{p_1}$. Letting $x = (\Sigma_{11})^{1/2} (b - \beta) (\Sigma_{22.1})^{-1/2}$, from (13) we have to evaluate

$${}_1F_1\left(\frac{n + 1}{2}; \frac{n}{2}; \frac{1}{2} \Phi^{-1}\right) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{((n + 1)/2)_{\kappa}}{(n/2)_{\kappa}} \frac{C_{\kappa}(\Phi^{-1}/2)}{k!} \quad \text{with } \Phi = I_{p_1} + xx' = (\Sigma_{11})^{1/2} \Theta (\Sigma_{11})^{1/2}. \quad (16)$$

Constantine's zonal polynomial $C_{\kappa}(\Phi^{-1}/2)$ is a linear form $C_{\kappa}(\Phi^{-1}/2) = \Xi_{\kappa} s(k)$ in the vector of monomials

$s(k) = (s_k, s_{k-1} s_1, s_{k-2} s_2, \dots, s_{k-2} s_1 s_1, s_{k-3} s_2 s_1, \dots, s_2 s_2 (s_1)^{k-4}, (s_1)^k)'$ with its elements $(s_{\ell_1} s_{\ell_2} \dots s_{\ell_k})$ corresponding to a partition $\lambda = (\ell_1, \ell_2, \dots, \ell_k)$ of k , ordered by shortest length.

and where

$$s_j = \text{sum of the } j\text{-th powers of the roots of } \frac{1}{2} \Phi^{-1}.$$

One root of Φ is equal to its determinant $|\Sigma_{11}||\Theta|$ and $p_1 - 1$ roots are equal to one. Since the roots of Φ^{-1} are the reciprocals of those of Φ , the sum of powers $s_j = \frac{1}{2^j}(p_1 - 1 + (|\Sigma_{11}||\Theta|)^{-j})$. Therefore the vector $s(k) = (p_\lambda(1/|\Sigma_{11}||\Theta|))$, where $p_\lambda(1/|\Sigma_{11}||\Theta|)$ is a polynomial of degree k in $1/|\Sigma_{11}||\Theta|$ and the hypergeometric function

$${}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2} \Phi^{-1}\right) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_x \frac{((n+1)/2)_k}{(n/2)_k} \sum_\lambda \Xi_{k\lambda} p_\lambda(1/|\Sigma_{11}||\Theta|) = \sum_{i=0}^{\infty} a_i \left(\frac{1}{|\Sigma_{11}||\Theta|}\right)^i$$

is a polynomial of infinite degree in $1/|\Sigma_{11}||\Theta|$. From (13) when $p_2 = 1$ and $M_1 M_1' (\Sigma_{11})^{-1} = I_{p_1}$ we get

$$f_s(b | p_1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n+1, M_2 = \beta M_1) = \sum_{i=0}^{\infty} \mu_i f_s(b | p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, n+1+2i, 0), \quad (17)$$

with $\mu_i = \exp(-p_1/2) \frac{((n+1-p_1)/2)_i}{((n+1)/2)_i} a_i$

ii). When $p_2 = 1$, p_1 and the means M are arbitrary, but still satisfying $M_1 = \beta M_2$, from (13) the hypergeometric function to be evaluated is

$${}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2} \Psi \Phi^{-1}\right) = {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2} Q' \Psi Q \Lambda^{-1}\right), \quad \text{with } \Psi = (\Sigma_{11})^{-1} M_1 M_1', \quad (18)$$

where $\Phi = Q' \Lambda Q$, Q orthonormal and Λ the diagonal matrix of characteristic roots of Φ , these roots being $p_1 - 1$ ones and the p_1 -th root is equal to $|\Sigma_{11}||\Theta|$. Therefore the characteristic roots ρ of $Q' \Psi Q \Lambda^{-1}$ satisfy

$$0 = |Q' \Psi Q \Lambda^{-1} - \rho I_{p_1}| = |\Psi Q \Lambda^{-1} Q' - \rho I_{p_1}| = \left| \Psi (I_{p_1} + (\frac{1}{|\Sigma_{11}||\Theta|} - 1) Q^{p_1} (Q^{p_1})') - \rho I_{p_1} \right| \quad (19)$$

where Q^{p_1} is the p_1 -th column of Q i.e. the p_1 -th eigen vector of $\Phi = I_{p_1} + xx'$, with $x = (\Sigma_{11})^{1/2} (b - \beta) (\Sigma_{22.1})^{-1/2}$. Since the latter is the vector $x/\sqrt{x'x}$, the principal minors of all orders

$$\left| \Psi_{11} + (\frac{1}{|\Sigma_{11}||\Theta|} - 1) \Psi_1 \frac{xx_1'}{x'x} \right| = (1 + (\frac{1}{|\Sigma_{11}||\Theta|} - 1) \frac{x_1' (\Psi_{11})^{-1} \Psi_1 x}{x'x}) |\Psi_{11}|, \quad \Phi = \begin{pmatrix} \Phi_{11} \\ \Phi_{21} \end{pmatrix} = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}, \quad (20)$$

are linear in $(|\Sigma_{11}||\Theta|)^{-1}$. When (19) is written as a polynomial in ρ , all coefficients are also linear in $(|\Sigma_{11}||\Theta|)^{-1}$. Therefore, the sum of powers $s_j = (\rho_1)^j + (\rho_2)^j + \dots + (\rho_{p_1})^j$, the vector of monomials $s(k)$ and the hypergeometric function (18) are polynomials in $(|\Sigma_{11}||\Theta|)^{-1}$. If we write

$${}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{1}{2} \Psi \Phi^{-1}\right) = \sum_{i=0}^{\infty} a_i(b) \left(\frac{1}{|\Sigma_{11}| |\Theta|}\right)^i \quad \text{then}$$

$$f_S(b | p_1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n+1, M_2 = \beta M_1) = \sum_{i=0}^{\infty} \mu_i(b) f_S(b | p_1 \times 1, \Sigma_{11}, \Sigma_{22.1}, n+1, 0) \quad (21)$$

$$\text{with } \mu_i(b) = \exp\left[-\frac{1}{2} \text{tr}(M_1 M_1' (\Sigma_{11})^{-1})\right] \frac{((n+1-p_1)/2)_i}{((n+1)/2)_i} a_i(b).$$

The density is again a weighted average of central student densities of higher degrees of freedom. But there is a vast difference between (17) and (21). Here the weights are random dependent on b since the principal minors in (20) depend on $x = (\Sigma_{11})^{-1}(b - \beta)(\Sigma_{22.1})^{-1/2}$. Whereas all marginal and conditional densities are known for the density (17), the marginal and conditional densities of (21) are not so simple and require evaluation of the expected value of ratios of quadratic forms that appear in (20).

4. The IVE b^{IV} is Multivariate non-central student : $M_2 = \gamma M_1, \gamma \neq \beta$.

If the regression model is $X_2 = \gamma' X_1 + U$, with $E(U) = 0$, the density of the least squares estimator when $M = \gamma' M$, $\gamma \neq \beta$. is obtained from the density $f(S^0)$ at (7) by integrating out S_{11} and $S_{22.1}$. The estimator is centered around $\beta = \gamma + (\Sigma_{11})^{-1} \Sigma_{1v}$ and is thus biased and inconsistent whenever $(\Sigma_{11})^{-1} \Sigma_{1v} \neq 0$.

In the econometric literature the theory holds that there is a $q \times n$ set of instrumental variables Z that are either fixed or independent of X and such that the means M are linear forms $M = \Pi Z$. Instead of the least squares estimator consider the instrumental variable estimator which is the least squares estimator after replacement of the variables by their expectations $XZ'(ZZ)^{-1}Z$. The $p \times p$ moment matrix $S = XZ'(ZZ)^{-1}ZX'$ has rank $\text{Min}(p, q)$, its density under the normality assumption is the noncentral Wishart $f_w(S | p, \Sigma, q, M)$ and $q - p_1$ is the degree of overidentification.

Starting from the noncentral Wishart $f_w(S | p, \Sigma, q, M)$ and the relation $M_2 = \gamma' M_1$, the distribution of the instrumental variable estimator b^{IV} is derivable in exactly the same way as above, only the degrees of freedom differ. Instead of (7) we have

$$\begin{aligned} f(S^0) &= f_w(S_{11} | p_1, \Theta^{-1}, q + p_2, 0) \quad f_w(S_{22.1} | p_2, \Sigma_{22.1}, q - p_1, 0) \\ &\quad \times f_S(b^{IV} | p_1 \times p_2, \Sigma_{11}, \Sigma_{22.1}, q + p_2, 0) \exp\left[-\frac{1}{2} \text{tr}(M_1' H(\gamma, \gamma) M_1)\right] \\ &\quad \times {}_0F_1\left(\frac{q}{2}; \frac{1}{4} M_1' [(\gamma - \beta)(\Sigma_{22.1})^{-1} S_{22.1} (\Sigma_{22.1})^{-1} (\gamma - \beta)' + H(\gamma, b^{IV}) S_{11} H(b^{IV}, \gamma)] M_1\right), \end{aligned} \quad (22)$$

$$\text{with } H(z_1, z_2) = (\Sigma_{11})^{-1} + (z_1 - \beta)(\Sigma_{22.1})^{-1}(z_2 - \beta)' \quad \text{and} \quad \Theta = H(b^{IV}, b^{IV}).$$

As an illustration, we consider the two Examples 2 and 3 that have been developed in the literature by a different method.

Example 2. The Exactly identifiable model, $p_2 = 1$, $q - p_1 = 0$.

When the number of instruments q is equal to the number p_1 of regressors, the equation is said to be exactly identifiable. In (22) the density of $S_{22.1}$ is the degenerate density (spike) with all the weight at zero. Consequently the hypergeometric function only depends on the second term in its argument. The reproductive property of hypergeometric functions [1] implies that

$$\int_{S_{11} > 0} \exp\left[-\frac{1}{2} \text{tr}(\Theta S_{11})\right] {}_0F_1\left(\frac{1}{4} M_1' H(\gamma, b^{IV}) S_{11} H(b^{IV}, \gamma) M_1\right) = \Gamma_{p_1}\left(\frac{p_1+1}{2}\right) |\Theta/2|^{-(p_1+1)/2} {}_1F_1\left(\frac{p_1+1}{2}; \frac{p_1}{2}; \frac{1}{2} M_1' H(\gamma, b^{IV}) \Theta^{-1} H(b^{IV}, \gamma) M_1\right) \quad (23)$$

Substituting this into (22) we get in the exactly identified model with $M_2 = \gamma' M_1$ that

$$f_S(b^{IV} | p_1 \times 1, \Sigma_{11}, \Sigma_{22.1}, p_1 + 1, M_2 = \gamma' M_1) = \exp\left[-\frac{1}{2} \text{tr}(M_1' H(\gamma, \gamma) M_1)\right] \times {}_1F_1\left(\frac{p_1+1}{2}; \frac{p_1}{2}; \frac{1}{2} M_1' H(\gamma, b^{IV}) \Theta^{-1} H(b^{IV}, \gamma) M_1\right) f_S(b^{IV} | p_1 \times 1, \Sigma_{11}, \Sigma_{22.1}, p_1 + 1, 0) \quad (24)$$

Sargan [12] stated the density (24) in a different but equivalent form.

In comparison to the density of the least squares estimator (13) when $M_2 = \beta' M_1$, (24) has the same factorization and the same parameters except that instead of n degrees of freedom there are only p_1 , $H(\gamma, \gamma)$ takes the place of $(\Sigma_{11})^{-1}$ in the exponential and $H(\gamma, b^{IV}) \Theta^{-1} H(b^{IV}, \gamma)$ replaces $(\Sigma_{11})^{-1} \Theta^{-1} (\Sigma_{11})^{-1}$ in the hypergeometric function.

In the univariate case with a single regressor the density of the instrumental variable estimator is

$$f_S(b^{IV} | 1 \times 1, \Sigma_{11}, \Sigma_{22.1}, 2, M_2 = \gamma M_1) = \sum_{k=0}^{\infty} \lambda_k(b^{IV}) f_S(b^{IV} | 1 \times 1, \Sigma_{11}, \Sigma_{22.1}, 2 + 2k, 0), \quad (25)$$

$$\text{with } \lambda_k(b^{IV}) = \frac{(M_1 M_1' \Sigma_{11} H(\gamma, b^{IV}) H(b^{IV}, \gamma))^k}{2^k k!} / \exp\left[\frac{1}{2} M_1 M_1' H(\gamma, \gamma)\right]$$

This is identical to the distribution (14) when $\gamma = \beta$. Whereas (25) again states the distribution as an average of central student densities, the weights are random and depend on the instrumental variable estimator b^{IV} .

Example 3. The single equation over-identified model, $p_2 = 1$, $q - p_1 > 0$.

Richardson [10] and Sawa [11] derived the distribution in this over-identified simple regression case using somewhat different methods. From (22) the solution can be arrived at by expanding binomially the powers of the sum of the two arguments in the hypergeometric function and collecting terms. Following our interpretation the density so obtained can be stated as

$$f_S(b^{IV} | 1 \times 1, \Sigma_{11}, \Sigma_{22.1}, q+1, M_2 = \gamma M_1) = \sum_{k=0}^{\infty} \mu_k(b^{IV}) f_S(b^{IV} | 1 \times 1, \Sigma_{11}, \Sigma_{22.1}, q+1+2k, 0),$$

$$\text{with } \mu_k(b^{IV}) = \frac{\Gamma((q-1+2k)/2)}{\Gamma((q-1)/2)} \lambda_k(b^{IV}) \sum_{j=0}^{\infty} \frac{\Gamma((q-1+2j)/2)}{\Gamma((q+2j+2k)/2)} \frac{(M_1 M_1' (\Sigma_{22.1})^{-1} (\gamma - \beta)^2)^j}{2^j j!}$$

with $\lambda_k(b^{IV})$ defined at (25).

For the multivariate cases, (22) is a polynomial and if the coefficients in the hypergeometric function are given online, powers can be integrated term by term, these being all expectations in central Wishart densities. It is feasible computer work, but the computational amount is very large indeed since parameter values and roots are to be changed continuously.

5. Concluding Remarks.

In 1964 James concluded his review paper [5] as follows: <For numerical evaluation of the probability density functions the power series expansions of the hypergeometric functions occurring in the distributions are of very limited value. If even one root is significant, it will take a very large number of terms of the series to give values of the likelihood function accurate enough to be of use. The importance of the preceding theory lies in the mathematical characterization of the functions involved in the distributions.>.

If this view is still valid today so are the multivariate densities (13), (22) and (25) interesting theoretically, but of little practical use. Partly James was expressing his frustration with the tedious calculations of the lower order zonal polynomials, in addition to the huge amount of computational work involved in calculating a density that depends continuously on many parameters.

In the Appendix we show an online computation method for zonal polynomials involved in bivariate and trivariate densities and there is absolutely no reason to share the opinion expressed by James about forty years ago. The theory can be made useful and it is important that it be practical because without it the claim that some approximation method is performing well is without proof.

References:

- [1] Constantine, A.G. (1963), Some Non-central Distribution Problems in Multivariate Analysis, *Annals of Mathematical Statistics* , **34**, 1270-1285.
- [2] David, F.N. and Kendall M.G.(1955), Tables of symmetric functions. Part V., *Biometrika* , **42**, 223-242.
- [3] James, A.T.(1955), The non-central Wishart Distribution, *Proc. Roy. Soc. (London)*, Ser. A, **229**, 364-366.
- [4] James, A.T.(1961), Zonal polynomials of the real positive definite symmetric matrices, *Ann. of Math.* **74**, 456-469.
- [5] James, A.T. (1964), Distributions of Matrix Variates and Latent Roots derived from Normal Samples,*Ann. Math. Stat.*, **35**, 475-501.
- [6] Kabe, D.G. (1968), On the Distribution of the Regression Coefficient Matrix of a Normal Distribution, *The Australian Journal of Statistics*, **10**, 21-23.
- [7] Kshirsagar, A.M.,(1961), Some Extensions of the Multivariate t-Distribution and the Multivariate Generalization of the Distribution of the Regression Coefficient, *Proceedings of the Cambridge Philosophical Society*, **57**, 80-85.
- [8] Parkhursts, A.M. and James, A.T. (1974), Zonal Polynomials of order 1 through 12. *Selected Tables in Mathematical Statistics* (Harter, H.L. and Owen, D.B. Eds.), 199-388, American Mathematical Society, Providence, R.I.
- [9] Phillips, P.C.B. (1980), The exact Distribution of Instrumental Variable Estimators containing n+1 Endogenous Variables, *Econometrica* , Vol. **48**, May, 861-878.
- [10] Richardson, D.H. (1968), The Exact Distribution of a Structural Coefficient Estimator, *Journal of the American Statistical Association* , **63**, 1214-1226.
- [11] Sawa, T. (1969), The Exact Sampling Distribution of Ordinary Least Squares and Two-stage Least Squares Estimators, *Journal of the American Statistical Association* , **64**, 923-937.
- [12] Sargan, J. D. (1976), Econometric Estimators and the Edgeworth Approximation, *Econometrica* , **44**, 421-428; and Erratum, *Econometrica* , **45** (1977), 272.
- [13] Wegge, L.L. (1971), The Finite Sampling Distribution of LSE with stochastic regressors, *Econometrica*, **39**, 241-251.
- [14] Wishart J. (1928), The generalized product moment distribution in samples from a normal multivariate population, *Biometrika* , **20**, 32-43.

On Line Zonal Polynomials: A Recursive Method

Below we show that the zonal polynomials of order k are linear combinations of their top (or lowest) zonal polynomial and of polynomials that are products of zonal polynomials of order lower than k . The linear relation is triangular and is obtainable through the standard orthogonalization method.

Zonal Polynomials.

We write zonal polynomials as a vector of linear forms in a matrix format. Each row lists the coefficients of one of the zonal polynomials. A partition

$$\kappa = (k_1, k_2, \dots, k_m), \quad 0 \leq k_m \leq k_{m-1} \leq \dots \leq k_2 \leq k_1, \quad \sum_{i=1}^m k_i = k,$$

of the integer k , has length j if k_j is the last positive integer in κ . To each partition corresponds a polynomial and we order the polynomials (rows) by the shortest length and lexicographically within the same length.

Let $s_j = \sum_{i=1}^m (\lambda_i(S))^j$ be the sum of the j th powers of the latent roots $\lambda_i(S)$ of a $m \times m$ matrix S . Given a partition κ of length j , define the sum of powers monomial $s_{\kappa_1} s_{\kappa_2} \dots s_{\kappa_j}$. Ordering the monomials by shortest length and lexicographically within the length of the corresponding partitions, define the vector of sum of powers monomials

$$s(k) = (s_k, s_{k-1}s_1, s_{k-2}s_2, \dots, s_{k-2}(s_1)^2, s_{k-3}s_2s_1, \dots, (s_2)^2(s_1)^{k-4}, s_2(s_1)^{k-2}, (s_1)^k)'$$

If K is the number of partitions of k , the vector of James' zonal polynomials $Z(k, S)$ of order k is the K -component vector $Z(k, S) = \Delta(k)s(k)$ with $K \times K$ $\Delta(k)$ satisfying orthogonality conditions with respect to the top row of $\Delta(k)$. Given a vector $x = (x_i)$, let $\text{diag } x$ be the diagonal matrix with x_i as i th diagonal element. James [5] shows

$$\Delta(k) (\text{diag } \Delta(k)_1)^{-1} \Delta(k)' \quad (1)$$

is a diagonal matrix, where $\Delta(k)_1$ is the top row of $\Delta(k)$ defined as

$$\Delta(k)_1 = (\Delta(k)_{1v}), \quad \Delta(k)_{1v} = 2^k k! / v_1! v_2! v_3! \dots 2^{v_1} 4^{v_2} 6^{v_3} \dots, \quad (2)$$

where, if $\Delta(k)_{1v}$ is the element in the j th column, $v = (v_1, v_2, \dots, v_k)$, $v_1 + 2v_2 + 3v_3 + \dots = k$, if the j -th partition of k , as ordered above, consists of v_1 ones, v_2 twos, v_3 threes, ..., and v_k the number k .

Example.

When $k=6$, the number of partitions $K=11$. With the vector of polynomials and sum of powers monomials

$$Z(6, \cdot) = (Z_6, Z_{51}, Z_{42}, Z_{33}, Z_{411}, Z_{321}, Z_{222}, Z_{2211}, Z_{21111}, Z_{111111})'$$

$$s(6) = (s_6, s_5s_1, s_4s_1, s_3s_3, s_4s_1s_1, s_3s_2s_1, s_2s_2s_2, (s_2s_1)^2, s_2(s_1)^4, (s_1)^6)'$$

the vector of zonal polynomials $Z(6, S) = \Delta(6)s(6)$, with

$$\Delta(6) = \begin{pmatrix} 3840 & 2304 & 1440 & 640 & 720 & 960 & 120 & 160 & 180 & 30 & 1 \\ -384 & 192 & -144 & -64 & 192 & 80 & -12 & 72 & 48 & 19 & 1 \\ -48 & -144 & 108 & -8 & -18 & 24 & 30 & 16 & 27 & 12 & 1 \\ -24 & -48 & -114 & 136 & -78 & 120 & -27 & -8 & 33 & 9 & 1 \\ 96 & -48 & -24 & 16 & 12 & -60 & -12 & 22 & -12 & 9 & 1 \\ 16 & 32 & -4 & -24 & -18 & 0 & -2 & -8 & 3 & 4 & 1 \\ 0 & 24 & -60 & 40 & 30 & -60 & 30 & -20 & 15 & 0 & 1 \\ -48 & 24 & 12 & 16 & -6 & 12 & 6 & 4 & -21 & 0 & 1 \\ -12 & -24 & 24 & 4 & 24 & 0 & -9 & -8 & 3 & -3 & 1 \\ 48 & -24 & -36 & -16 & -6 & 20 & 6 & 12 & 3 & -8 & 1 \\ -120 & 144 & 90 & 40 & -90 & -120 & -15 & 40 & 45 & -15 & 1 \end{pmatrix}$$

James Calculation Method.

James [5] constructs an indirect calculation method, involving elementary symmetric functions $a(k)$ instead of the sum of powers monomials $s(k)$. Since the orthogonal relations (1) are in the space of $s(k)$, conversion tables relating the functions $a(k)$ to the monomials $s(k)$ are needed. David and Kendall [2] have published conversion tables for $k=1,..,12$. Correspondingly James tabulates zonal polynomials up to order 12 [8]. If a computer code is available relating $a(k)$ to $s(k)$, the need of tables would be eliminated. There is however still the inconvenience that under James' method, polynomials are built upwards from below, beginning with the last row and constructing the next higher up row by the standard orthogonalization method.

A recursive method.

In practice, only the polynomials corresponding to partitions up to length m will be needed, since the polynomials of length more than m are identically zero. In the alternative method introduced here there is no need to derive all polynomials of all lengths. Using (2), we start with the top polynomial of length one and calculate the next lower rows by orthogonalization of an auxiliary matrix $\Psi(k)$ of linearly independent vectors that form a basis for $\Delta(k)$. This method is fully self-contained and direct in deriving the zonal polynomials of higher rank and shorter length. The method proceeds as follows:

1. With $\kappa = (k_1, k_2, \dots, k_k)$ a partition of k , construct the row Ψ_κ such that

$$\Psi_\kappa s(k) = \Delta_{\{k_1\}} s(k_1) \Delta_{\{k_2, \dots, k_k\}} s(k - k_1) \quad (3)$$

The $K \times K$ matrix $\Psi(k) = (\Psi_\kappa)$ with κ in the order established above.

Notice that $\Delta_{\{k_1\}}$ is the top row in $\Delta(k_1)$ and $\Delta_{\{k_2, \dots, k_k\}}$ is one of the rows in $\Delta(k - k_1)$. Also the product of an element of $s(k_1)$ with an element of $s(k - k_1)$ is an element in $s(k)$. The last column of $\Psi(k)$ is a column of ones.

Computation Results.

We have programmed the recursive method described above in FORTRAN and run it on a Power Macintosh 7100/80AV to calculate the zonal polynomials of orders up to 23 and lengths up to 3. The storage requirements are high since all polynomials of lower order and length 2 are stored. Also when $k=23$, a 56–1255 $B_l(k)$ matrix is needed and an array of 60929 entries to store the polynomials of length up to two.

We are encouraged however that the zonal polynomials of length 2 of all orders can be produced online on a small PC, since then only the top polynomial coefficients $\Delta(k)_1$ have to be stored and in addition we have verified that the elements of the $B(k)$ -matrices have the simple elegant algebraic form

$$B(k) = (b_{k_1 k_2; \ell_1 \ell_2}) = \left(\frac{k_2!}{(k_2 - \ell_2)! \ell_2!} \frac{(-1/2)_{k_2 - \ell_2}}{(k_1 - \ell_2)_{k_2 - \ell_2}} \frac{(k_1 - \ell_2 + 3/2)_{\ell_2}}{(\ell_1 - \ell_2 + 1)_{\ell_2}} \right), \quad 0 \leq \ell_2 \leq k_2,$$
$$= 0, \quad k_2 < \ell_2.$$

We also obtained the formula for the zonal polynomials of three parts. As a consequence online calculations in univariate, bivariate and trivariate applications are resolved.