

MARGINAL DENSITIES OF INSTRUMENTAL VARIABLE ESTIMATORS IN THE GENERAL SINGLE EQUATION CASE

P. C. B. Phillips

ABSTRACT

A method of extracting marginal density approximations using the multivariate version of the Laplace formula is given and applied to instrumental variable estimators. Some leading exact distributions are derived for the general single equation case which lead to computable formulae and generalize all known results for marginal densities. These results are related to earlier work by Basmann (1963), Kabe (1964), and Phillips (1980b). Some general issues bearing on the current development of small-sample theory and its application in empirical work are discussed in the introduction to the article.

Advances in Econometrics, Volume 2, pages 1-24

Copyright © 1983 by JAI PRESS INC.

All rights of reproduction in any form reserved.

ISBN: 0-89232-183-0

I. INTRODUCTION

The general problem of extracting finite sample distributions in econometrics has attracted attention since the early 1960s. One reason for the interest has been the widely acknowledged shortage of results in the area and the heavy reliance empirical investigators have, therefore, reluctantly had to place on asymptotic distribution theory in estimation and inference. That such a heavy reliance on asymptotic theory can lead to serious problems of bias and low levels of inferential accuracy in small sample situations is well understood, particularly by economists who have had to deal with short data series in empirical work. Moreover, some of the earliest results on small-sample distributions, particularly in time series models [for example, the work of Hurwicz (1950), which clearly illustrated the substantial small sample bias of the least squares estimator in the first order autoregression], must have made many investigators in the profession uneasy about setting the foundation for estimation and inferential procedures on asymptotic theory alone. Nevertheless, these procedures have out of necessity become firmly established, and the conventional coding of asymptotic statistics in computer regression outputs has helped to entrench the practice of a sole reliance on asymptotic theory in empirical work.

Since the early papers 20 years ago (Basmann, 1961; Bergstrom, 1962; Kabe, 1963, 1964), there has been a continuing and growing literature on small-sample distribution theory in econometrics. Three major schools of research emerged in the 1970s associated with the names of Professor T. W. Anderson at Stanford University, Professor R. L. Basmann at Texas A and M University, and Professor J. D. Sargan at the London School of Economics. The research work of these schools has very largely been complementary in extending the frontiers of knowledge in this field and in stimulating the interest of new research workers. Although the ultimate objective of this research has been to relieve the empirical worker from the heavy reliance he has had to place on asymptotic theory, as yet there has been no substantial payoff to this research in terms of applied econometric practice. This situation is most likely to change dramatically during the 1980s. In part, this is because the rather specialized results of the early research have recently given way to general theories and a powerful technical machinery that will make it easier to transmit results and methods to the applied econometrician in the precise setting of the model and the data set with which he is working. And, in part, this is because improvements in computing now make it feasible to incorporate into existing regression software subroutines that will provide the essential vehicle for this transmission.

Two parallel current developments in the subject are an integral part of this process. The first of these is concerned with the derivation of direct approximations to the sampling distributions of interest in an applied study. These approximations can then be utilized in the decisions that have to be made by an investigator concerning, for instance, the choice of an estimator or the specification of a critical region in a statistical test. Techniques that offer most promise in this regard are the Edgeworth approximation, whose use has been explored in general cases by Sargan (1976) and Phillips (1977, 1980a), and the modified Padé approximant introduced in Phillips (1982).

The second relevant development involves advancements in the mathematical task of extracting the form of exact sampling distributions in econometrics. In the context of simultaneous equations, the published literature has so far concentrated on the sampling distributions of estimators and test statistics in single structural equations involving only two, or at most three, endogenous variables. Recent theoretical work by Phillips and Rhodes has extended much of this work to the general case. In particular, Phillips (1980b) extracted the exact density function of the instrumental variable (IV) estimator in the most general case of a structural equation with $n + 1$ endogenous variables and an arbitrary number of degrees of overidentification. In the same setting, Rhodes (1981) found the exact density of the likelihood ratio identifiability test statistic. Work that the author currently has underway should extend these general results to the k -class estimators and limited-information maximum likelihood (LIML). However, in spite of their generality, these results suffer a major handicap in computational work. The analytical methods used in these articles rely on the manipulation of matrix argument higher transcendental functions, and the final expressions for the density functions that have been obtained involve multiple infinite series in terms of zonal-type polynomials.

Zonal polynomials were introduced by James (1961) and can be represented as symmetric homogeneous polynomials in the latent roots of their matrix arguments. Extensions of these polynomials to include similar polynomials in two or more matrix arguments have been made by Davis (1979, 1980) and Chikuse (1981) and are also relevant in certain econometric applications (see Appendix B of Phillips, 1980b). These polynomials we will refer to under the generic name zonal-type polynomials. Although the series representations that involve these polynomials are very convenient mathematically, they present enormous difficulties in numerical work. This is, in part, due to the fact that no general formulae for the zonal-type polynomials are known. However, certain algorithms for the computation of the zonal polynomials themselves are available [see, in

particular, James (1968) and McLaren (1976)],¹ and a complete computer program for the evaluation of the coefficients of these polynomials has been developed and made available by Nagel (1981). This is an important development and will in due course undoubtedly affect our present rather limited ability to numerically compute and readily interpret multiple series representations of probability density functions. Unfortunately, the availability of tabulations and algorithms for the zonal-type polynomials will cover only part of the computational difficulty. As noted by Muirhead (1978), the series that involves these polynomials often converge very slowly. This problem arises particularly when the polynomials have large arguments (large latent roots) and it becomes necessary to work deeply into the higher terms of the series in order to achieve convergence. This in turn raises additional problems of underflow and overflow in the computer evaluations of the coefficients in the series and the polynomials themselves. To take as a simple example the case of the exact density of the IV estimator in the two endogenous variable case, the author has found that in a crude summation of the doubly infinite series for the density a thousand or more terms seem to be necessary² to achieve adequate convergence when the true coefficient parameter is greater than 5 and the concentration parameter greater than 10. These are not in any way unrealistic values, and the problems increase with the size of the coefficient and concentration parameter. Expressed as a single series involving the ${}_1F_1$ function of a scalar argument, we find that computation of the series requires computation of the ${}_1F_1$ function for a scalar argument greater than 225. Use of the conventional asymptotic expansion of the ${}_1F_1$ function [which is normally recommended when the argument is greater than 10; see Slater (1965)] fails here because one of the parameters of the ${}_1F_1$ function grows as we enter more deeply into the series and the series itself no longer converges. Some special methods for dealing with these difficulties in this case have been developed in programming work for the author by Sidnie Feit and will be reported elsewhere. Undoubtedly, the additional problems we have encountered in this example quickly become much worse as the dimension of the argument matrices in the special functions and the zonal polynomials increases and as we need to make use of the more general zonal-type polynomials.

For direct computational work in the case of the IV estimator when there are more than two endogenous variables in the structural equation, the problems reported in the previous paragraph were overcome in Phillips (1980b) by extracting an asymptotic expansion of the exact joint density of the vector coefficient estimator. The leading term of this expansion has an error of $O(T^{-1})$ where T is the sample size, and in the univariate (two endogenous variable) case the resulting approximation is the saddle-

point approximation derived using other methods by Holly and Phillips (1979). As shown in this latter article, the approximation gives high accuracy for some plausible values of the parameters throughout a wide domain of the distribution, including the tails.

The rather specialized early research in small-sample theory naturally enough focused on problems that were essentially univariate in character in the sense that the statistics under analysis were scalar rather than vector random variates. Hence, in the context of a structural equation with two endogenous variables, attention concentrated on the sampling distributions of (1) estimators of the unknown coefficient of one of the endogenous variables, (2) structural variance estimators, and (3) t -ratio type test statistics. As the analytic theory has developed to encompass the more general cases discussed in previous paragraphs, the final problem has, at least in the important case (1) above, become multivariate³ in character. For example, in the work by Phillips (1980b) on the IV estimator, the structural equation contains $n + 1$ endogenous variables, there are n unknown structural coefficients of these variables, and the final probability density function (pdf) of the estimator is a joint density in n dimensions.

In this multivariate setting, an important and outstanding problem is the characterization of the joint density in such a way that the final results can intelligently be used in practice. This problem bears an immediate resemblance to that of characterizing multidimensional Bayesian posterior distributions. Some of the current work in the latter area has taken the direction of numerical computation of posterior moments by Monte Carlo methods (see, for instance, Kloek and Van Dijk, 1978). Direct application of these techniques is hampered by the difficulties typically involved in the computation of the joint densities that have been described earlier (although approximations to the joint density functions could be used before applying this method). Moreover, Monte Carlo methods have concentrated on the computation of low order posterior moments. In many cases it will be preferable to work with the marginal density functions directly, since attention is often focused on individual parameters or coefficients rather than a group of parameters jointly.⁴ In principle, the marginal densities can be obtained by analytic or numerical integration over the space of the auxiliary variables and for the domain of values for which the marginals are required. In practice, analytic integration currently presents severe technical difficulties for all but the simplest cases (and with the exception of the analysis of the leading terms in the density expansions); and numerical integration methods encounter dimensionality problems as well as/or computational difficulties in the joint density series representations described earlier. It is interesting to note that of the two significant parallel developments in small sample

theory that were discussed earlier, the first of these yields marginal density approximations directly. This is an attractive feature of both the Edgeworth and the modified Padé approximants.

When the starting point in our analysis is the exact joint pdf of a group of parameters and we wish to sharpen our focus and characterize the sampling behavior of an individual parameter estimator, an alternative approach is required. This article presents a solution to this problem, a solution that involves the use of the multivariate version of Laplace's method [see, for example, Chapter 8 of Bleistein and Handelsman (1976)] to reduce the multidimensional integrals that define the marginals. This approach should be quite generally applicable in both sampling theoretic and Bayesian problems and seems to present no computational problems even in high-dimensional cases.

Section II of the paper will outline the essential features of the method, and an application to the marginal densities of the IV estimator in the general single equation case will be presented in Section III. This application enables us to assess, *inter alia*, the effect on the density of an individual coefficient estimator of an increase in the number of endogenous variables in the model, *ceteris paribus*. Some graphs that display this and other effects are given. Some leading exact results will be given in Section IV; these results lead to computable formulae and generalize all presently known results for exact marginal densities in single structural equations. Concluding remarks will be made in Section V.

II. ASYMPTOTIC EXPANSION OF THE MULTIPLE INTEGRAL DEFINING A MARGINAL DENSITY

To set up a general framework, we assume that a model is specified that uniquely determines the joint probability distribution of the system's endogenous variables $\{y_1, \dots, y_T\}$ condition on certain fixed exogenous variables $\{x_1, \dots, x_T\}$, where T is the sample size. If, as is usual, this distribution is absolutely continuous, it can be represented by its probability density function (pdf), which will depend in general on an unknown vector of parameters θ . Thus, we write $\text{pdf}(y|x; \theta)$ where $y' = (y'_1, \dots, y'_T)$ and $x' = (x'_1, \dots, x'_T)$. Estimation of θ or a subvector of θ leads to a function of the available data, which we will write as the n -vector $\theta_T = \theta_T(y, x)$. Since most econometric estimators and test statistics are relatively simple functions of the sample moments of the data, which we will denote by m , we may also conveniently write $\theta_T = \theta_T^*(m)$. Frequently, these functions will just involve rational functions of the first and second sample moments of the data. In the more complicated cases, θ_T will be determined by a system of estimating equations, such

as $H_T(\boldsymbol{\theta}_T, m) = 0$, whose elements are usually rational functions of the sample moments. In any case, the representation $\boldsymbol{\theta}_T = \boldsymbol{\theta}_T^*(m)$ will be general enough for a large class of applications.

Our problem is to characterize the distribution of individual components of $\boldsymbol{\theta}_T$. In many of the relevant econometric cases, we can write down directly the pdf of the sample moments, namely, $\text{pdf}(m)$, using established results from multivariate distribution theory. Otherwise, $\text{pdf}(m)$ must be obtained by transformation and analytic integration of $\text{pdf}(\mathbf{y})$. The next step is to find a suitable set of auxiliary variates \mathbf{b} for which the transformation $m \rightarrow (\boldsymbol{\theta}_T, \mathbf{b})$ is 1:1 and then the density of $\boldsymbol{\theta}_T$ is given by the integral

$$\text{pdf}(\mathbf{r}) = \int_B \text{pdf}(m) \left| \frac{\partial m}{\partial(\mathbf{r}, \mathbf{b})} \right| d\mathbf{b} \quad (1)$$

where B is the space of definition of the \mathbf{b} variates and we use \mathbf{r} to denote $\boldsymbol{\theta}_T$ in the density. The marginal density of an individual component of $\boldsymbol{\theta}_T$ —say, θ_{Ti} —follows by further integration

$$\text{pdf}(\mathbf{r}_i) = \int_{R_i} \int_B \text{pdf}(m) \left| \frac{\partial m}{\partial(\mathbf{r}, \mathbf{b})} \right| d\mathbf{b} d\mathbf{r}_{(i)} \quad (2)$$

where $\mathbf{r}_{(i)}$ is the vector of redundant variables in \mathbf{r} and R_i is their space of definition. In (2) we use \mathbf{r}_i to represent $\boldsymbol{\theta}_{Ti}$.

Although the integral defining (2) can be obtained analytically in some leading general cases for simultaneous equations estimators (we will give some examples in Section IV), the complexity of the representation of the joint density $\text{pdf}(\mathbf{r})$ derived from (1) will normally present a severe obstacle to this step. Typically, $\text{pdf}(\mathbf{r})$ has an analytic representation as a multiple infinite series involving zonal polynomials of argument matrices that are themselves rational functions of the elements of \mathbf{r} . In the absence of analytic formulae for these polynomials in terms of the elements of their argument matrices, the task of integrating out the surplus elements of \mathbf{r} to extract the marginals $\text{pdf}(\mathbf{r}_i)$ will be possible only in the simplest cases.

An alternative approach, which simplifies the preceding task and which should lead to good asymptotic approximations in many cases, is based on the multivariate extension of Laplace's method applied to the multiple integrals (1) and (2). A full discussion of the method in the context of multidimensional integrals is given in Chapter 8 of Bleistein and Handelsman (1976). The method is applicable to integrals of the form

$$I(\lambda) = \int_D \exp\{\lambda\phi(\mathbf{x})\}g(\mathbf{x})d\mathbf{x}, \quad \mathbf{x} = (x_1, \dots, x_N) \quad (3)$$

where D is a simply connected domain in N -dimensional \mathbf{x} -space. If $\phi(\mathbf{x})$ and $g(\mathbf{x})$ are continuously differentiable to second order, if ϕ obtains an interior absolute maximum at \mathbf{x}^* in D , and if the Hessian $\partial^2\phi(\mathbf{x}^*)/\partial\mathbf{x}\partial\mathbf{x}'$ is negative definite, then [equation (8.3.52) of Bleistein and Handelsman (1976)]

$$I(\lambda) \sim (2\pi/\lambda)^{N/2} \exp\{\lambda\phi(\mathbf{x}^*)\}g(\mathbf{x}^*)\{\det[-\partial^2\phi(\mathbf{x}^*)/\partial\mathbf{x}\partial\mathbf{x}']\}^{-1/2} \quad (4)$$

in the sense that $A \sim B$ if $A/B \rightarrow 1$ as $\lambda \rightarrow \infty$. The error on this approximation is of $O[\exp\{\lambda\phi(\mathbf{x}^*)\}\lambda^{-(N+2)/2}]$ as $\lambda \rightarrow \infty$. The right-hand side of (4) is, in fact, the leading term in an asymptotic expansion of $I(\lambda)$ as $\lambda \rightarrow \infty$ provided the functions $\phi(\mathbf{x})$ and $g(\mathbf{x})$ satisfy sufficient regularity conditions. Equation (8.3.50) of Bleistein and Handelsman (1976) gives this expansion explicitly. It is worth pointing out that (3) holds when \mathbf{x}^* is an interior point of D and somewhat different formulas apply when the maximum of $\phi(\mathbf{x})$ is attained on the boundary of D . Intuitively, the asymptotic approximation (4) is based on the notion that as λ grows large the main contribution to the value of the integral comes from integrating in a small neighborhood of the point \mathbf{x}^* where $\phi(\mathbf{x})$ attains its absolute maximum.

In applying (4) to the multiple integrals that define $\text{pdf}(\mathbf{r})$ and $\text{pdf}(\mathbf{r},)$, we first need to represent the integrand in each case in a form that corresponds to that of (3). To fix ideas, we assume that (1) has the alternative representation

$$\text{pdf}(\mathbf{r}) = h(\mathbf{r}) \int_B \exp\{\lambda(T)\phi(\mathbf{r},\mathbf{b})\}g(\mathbf{r},\mathbf{b})d\mathbf{b} \quad (5)$$

for sufficiently smooth functions h , ϕ , and g and where $\lambda(T) \rightarrow \infty$ as $T \rightarrow \infty$.⁵ If $\phi(\mathbf{r},\mathbf{b})$ attains an interior absolute maximum at $\mathbf{b}^* = \mathbf{b}^*(\mathbf{r})$ in the interior of B , we have directly the asymptotic representation

$$\text{pdf}(\mathbf{r}) = [2\pi/\lambda(T)]^{N/2} h(\mathbf{r})\exp\{\lambda(T)\phi^*(\mathbf{r})\}g^*(\mathbf{r})[\Delta(\mathbf{r})]^{-1/2} \quad (6)$$

where

$$\phi^*(\mathbf{r}) = \phi[\mathbf{r},\mathbf{b}^*(\mathbf{r})]$$

$$g^*(\mathbf{r}) = g[\mathbf{r},\mathbf{b}^*(\mathbf{r})]$$

and

$$\Delta(\mathbf{r}) = \det\{-\partial^2\phi[\mathbf{r},\mathbf{b}^*(\mathbf{r})]/\partial\mathbf{b}\partial\mathbf{b}'\}.$$

To approximate the marginal density (2), we apply the method again, this time to the integral (2) giving

$$\text{pdf}(\mathbf{r}_t) = [2\pi/\lambda(T)]^{(N+n-1)/2} h(\mathbf{r}_t, \mathbf{r}_{(t)}^*) \exp\{\lambda(T)\phi^*(\mathbf{r}_t, \mathbf{r}_{(t)}^*)\} \cdot g^*(\mathbf{r}_t, \mathbf{r}_{(t)}^*) [\Delta(\mathbf{r}_t, \mathbf{r}_{(t)}^*)]^{-1/2} [\Delta_t(\mathbf{r}_t)]^{-1/2} \quad (7)$$

where

$$\Delta_t(\mathbf{r}_t) = \det\{-\partial^2\phi^*[\mathbf{r}_t, \mathbf{r}_{(t)}^*(\mathbf{r}_t)]/\partial\mathbf{r}_{(t)}\partial\mathbf{r}_{(t)}'\}$$

and where $\phi^*(\mathbf{r}) = \phi^*(\mathbf{r}_t, \mathbf{r}_{(t)}^*)$ attains an absolute maximum in the interior of R_t , given \mathbf{r}_t , at $\mathbf{r}_{(t)}^* = \mathbf{r}_{(t)}^*(\mathbf{r}_t)$. The dimension of the vector $\mathbf{r}_{(t)}^*$ here is $n - 1$.

III. AN APPLICATION TO INSTRUMENTAL VARIABLE ESTIMATORS

As in Phillips (1980b), we work with the structural equation

$$\mathbf{y}_1 = \mathbf{Y}_2\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\gamma} + \mathbf{u} \quad (8)$$

where \mathbf{y}_1 ($T \times 1$) and \mathbf{Y}_2 ($T \times n$) are an observation vector and matrix, respectively, of $n + 1$ included endogenous variables; \mathbf{Z}_1 is a $T \times K_1$ matrix of included exogenous variables; and \mathbf{u} is a random disturbance vector. The reduced form of (8) is given by

$$[\mathbf{y}_1; \mathbf{Y}_2] = [\mathbf{Z}_1; \mathbf{Z}_2] \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} + [\mathbf{v}_1; \mathbf{V}_2] = \mathbf{z}\boldsymbol{\Pi} + \mathbf{V} \quad (9)$$

where \mathbf{Z}_2 is a $T \times K_2$ matrix of exogenous variables excluded from (8). The rows of the reduced form disturbance matrix \mathbf{V} are assumed to be independent, identically distributed, normal random vectors. We assume that the usual standardizing transformations (Basmann, 1963, 1974) have been carried out, so that the covariance matrix of rows of \mathbf{V} is the identity matrix and $T^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{I}_K$ where $K = K_1 + K_2$ and $\mathbf{Z} = [\mathbf{Z}_1; \mathbf{Z}_2]$. We also assume that $K_2 \geq n$ and the matrix $\boldsymbol{\Pi}_{22}$ ($K_2 \times n$) in (9) has full rank, so that (8) is identified, although this latter assumption will be relaxed in our discussion of some leading cases in Section IV. We let $H = [\mathbf{Z}_1; \mathbf{Z}_3]$, where \mathbf{Z}_3 ($T \times K_3$) is a submatrix of \mathbf{Z}_2 and $K_2 \geq n$, be a matrix of IVs to be used in the estimation of (8). The IV estimator of the parameter vector $\boldsymbol{\beta}$ in (8) is then

$$\boldsymbol{\beta}_{IV} = (\mathbf{Y}_2'\mathbf{M}_H\mathbf{Y}_2)^{-1}(\mathbf{Y}_2'\mathbf{M}_H\mathbf{y}_1)$$

where $\mathbf{M}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}' - \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'$, and under orthogonality this reduces to $\boldsymbol{\beta}_{IV} = (\mathbf{Y}_2'\mathbf{Z}_3\mathbf{Z}_3'\mathbf{Y}_2)^{-1}(\mathbf{Y}_2'\mathbf{Z}_3\mathbf{Z}_3'\mathbf{y}_1)$.

The exact joint probability density of $\boldsymbol{\beta}_{IV}$ obtained by Phillips (1980b)

is given by

$$\begin{aligned} \text{pdf}(\mathbf{r}) &= \frac{\text{etr}\{- (T/2)(\mathbf{I} + \boldsymbol{\beta}\boldsymbol{\beta}')\bar{\boldsymbol{\Pi}}_{22}'\bar{\boldsymbol{\Pi}}_{22}\}\Gamma_n[(L+n+1)/2]}{\pi^{n/2}[\det(\mathbf{I} + \mathbf{r}\mathbf{r}')]^{(L+n+1)/2}} \\ &\cdot \sum_{j=0}^{\infty} \frac{(L/2)^j}{j!\Gamma_n[(L+n)/2+j]} [\{ (T/2)\boldsymbol{\beta}'\bar{\boldsymbol{\Pi}}_{22}'[\text{adj}(\partial/\partial\mathbf{W})] \\ &\bar{\boldsymbol{\Pi}}_{22}\boldsymbol{\beta}'\}^j [\det(\mathbf{I} + \mathbf{W})]^{(L-1)/2+j} \\ &\cdot {}_1F_1\{(L+n+1)/2, (L+n)/2+j; (T/2)(\mathbf{I} + \mathbf{W}) \\ &\bar{\boldsymbol{\Pi}}_{22}(\mathbf{I} + \boldsymbol{\beta}\mathbf{r}')(\mathbf{I} + \mathbf{r}\mathbf{r}')^{-1}(\mathbf{I} + \mathbf{r}\boldsymbol{\beta}')\bar{\boldsymbol{\Pi}}_{22}'\}_{\mathbf{W}=0} \end{aligned} \quad (10)$$

where $L = K_3 - n$ (the number of surplus instruments used in the estimation of $\boldsymbol{\beta}$) and $\bar{\boldsymbol{\Pi}}_{22}$ is an $n \times n$ matrix [nonsingular when $\boldsymbol{\Pi}_{22}$ in (9) is of full rank] defined by the equation $\boldsymbol{\Pi}_{22}\bar{\boldsymbol{\Pi}}_{22} = \bar{\boldsymbol{\Pi}}_{22}'\boldsymbol{\Pi}_{22}$. In (10) as in (1), we use \mathbf{r} to represent the estimator (in this case $\boldsymbol{\beta}_{IV}$), and \mathbf{W} is a matrix of auxiliary variables.

The matrix argument in the ${}_1F_1$ function that appears in the preceding expression for the exact density (10) has elements that are rational functions in the components of \mathbf{r} . The multiple series representation of (10) will similarly involve zonal polynomials in the same matrix argument; and as discussed in the previous section, this feature of the problem combined with the absence of analytic formulae for the polynomials in terms of the elements of the argument matrix makes the analytic determination of marginal densities difficult except for specialized leading cases.

Using the asymptotic representation of the ${}_1F_1$ function, namely, as $T \rightarrow \infty$ and for a nonsingular matrix \mathbf{R}

$${}_1F_1(a, b; \mathbf{TR}) = [\Gamma_n(b)/\Gamma_n(a)]\text{etr}(\mathbf{TR})(\det \mathbf{TR})^{a-b}[1 + O(T^{-1})]$$

(Constantine and Muirhead (1976), Theorem 3.2) in (10), we obtain the following asymptotic approximation of the joint density of $\boldsymbol{\beta}_{IV}$

$$\begin{aligned} &\text{pdf}(\mathbf{r}) \\ &T^{n/2}\text{etr} - \frac{T}{2} \left\{ \frac{\mathbf{M}(\mathbf{r} - \boldsymbol{\beta})(\mathbf{r} - \boldsymbol{\beta}')}{1 + \mathbf{r}'\mathbf{r}} \right\} (\det \mathbf{M})^{1/2} \\ &\sim \frac{(1 + \boldsymbol{\beta}'\mathbf{r})^{L+1}}{2^{n/2}\pi^{n/2}(1 + \mathbf{r}'\mathbf{r})^{(L+n+2)/2}} \cdot \frac{1}{(1 + 2\boldsymbol{\beta}'\mathbf{r} - \boldsymbol{\beta}'\boldsymbol{\beta})^{L/2}} \end{aligned} \quad (11)$$

[see Phillips (1980b) equation (15)] where $\mathbf{M} = \bar{\boldsymbol{\Pi}}_{22}'\bar{\boldsymbol{\Pi}}_{22}$ and which holds with a relative error of $O(T^{-1})$ as $T \rightarrow \infty$ for values of \mathbf{r} in the domain defined by $1 + 2\boldsymbol{\beta}'\mathbf{r} - \boldsymbol{\beta}'\boldsymbol{\beta} > 0$. Alternatively, (11) can be extracted

using an asymptotic expansion of a multiple integral similar to (5) but of the Fourier type (see Note 5).

We can now apply Laplace's method directly to extract marginal densities from (11). Without loss of generality, our attention will concentrate on $\text{pdf}(\mathbf{r}_1)$. We start by partitioning the matrix \mathbf{M} and vectors $\boldsymbol{\beta}$ and \mathbf{r} as

$$\mathbf{M} = \begin{bmatrix} m_{11} & \mathbf{m}'_{21} \\ \mathbf{m}_{21} & \mathbf{M}_{22} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \text{ and } \mathbf{r} = \begin{bmatrix} r_1 \\ \mathbf{r}_2 \end{bmatrix}$$

where m_{11} , β_1 , and r_1 are scalars. We define

$$\bar{m}_{11} = m_{11} - \mathbf{m}'_{21} \mathbf{M}_{22}^{-1} \mathbf{m}_{21}$$

$$\bar{\mathbf{r}}_2 = \bar{\mathbf{r}}_2(r_1) = \boldsymbol{\beta}_2 - (r_1 - \beta_1) \mathbf{M}_{22}^{-1} \mathbf{m}_{21}$$

$$\phi(r_1, \mathbf{r}_2) = \frac{(r_1 - \beta_1)^2 \bar{m}_{11} + (\mathbf{r}_2 - \bar{\mathbf{r}}_2)' \mathbf{M}_{22} (\mathbf{r}_2 - \bar{\mathbf{r}}_2)}{1 + r_1^2 + \mathbf{r}'_2 \mathbf{r}_2}$$

$$g(r_1, \mathbf{r}_2) = \frac{(1 + \beta_1 r_1 + \boldsymbol{\beta}'_2 \mathbf{r}_2)^{L+1}}{(1 + r_1^2 + \mathbf{r}'_2 \mathbf{r}_2)^{(L+n+2)/2} (1 + 2\beta_1 r_1 + 2\boldsymbol{\beta}'_2 \mathbf{r}_2 - \boldsymbol{\beta}' \boldsymbol{\beta})^{L/2}}.$$

By use of (4) in the $N = n - 1$ dimensional integral defining $\text{pdf}(r_1)$, we obtain the following asymptotic approximation to the marginal density

$$\text{pdf}(r_1) \sim \left(\frac{T}{2\pi} \right)^{1/2} \frac{(\det \mathbf{M})^{1/2}}{[\det \mathbf{H}(r_1, \mathbf{r}_2^*(r_1))]^{1/2}} g[r_1, \mathbf{r}_2^*(r_1)] \exp \left\{ -\frac{T}{2} \phi[r_1, \mathbf{r}_2^*(r_1)] \right\}$$

where

$$\begin{aligned} \mathbf{H}(r_1, \mathbf{r}_2) &= \partial^2 [(1/2)\phi(r_1, \mathbf{r}_2)] / \partial \mathbf{r}_2 \partial \mathbf{r}'_2 \\ &= [\alpha_1(r_1, \mathbf{r}_2) \mathbf{I} + \alpha_2(r_1, \mathbf{r}_2) \mathbf{r}_2 (\mathbf{r}_2 - \bar{\mathbf{r}}_2)'] \mathbf{M}_{22} \\ &\quad + \alpha_3(r_1, \mathbf{r}_2) \mathbf{I} + \alpha_4(r_1, \mathbf{r}_2) \mathbf{r}_2 \mathbf{r}'_2 \end{aligned}$$

and the α_i are scalar functions given by

$$\alpha_1(r_1, \mathbf{r}_2) = (1 + r_1^2 + \mathbf{r}'_2 \mathbf{r}_2)^{-1}$$

$$\alpha_2(r_1, \mathbf{r}_2) = -4(1 + r_1^2 + \mathbf{r}'_2 \mathbf{r}_2)^{-2}$$

$$\alpha_3(r_1, \mathbf{r}_2) = -(1 + r_1^2 + \mathbf{r}'_2 \mathbf{r}_2)^{-2} [(r_1 - \beta_1)^2 \bar{m}_{11} + (\mathbf{r}_2 - \bar{\mathbf{r}}_2)' \mathbf{M}_{22} (\mathbf{r}_2 - \bar{\mathbf{r}}_2)]$$

$$\alpha_4(r_1, \mathbf{r}_2) = 4(1 + r_1^2 + \mathbf{r}'_2 \mathbf{r}_2)^{-3} [(r_1 - \beta_1)^2 \bar{m}_{11} + (\mathbf{r}_2 - \bar{\mathbf{r}}_2)' \mathbf{M}_{22} (\mathbf{r}_2 - \bar{\mathbf{r}}_2)]$$

In (12), $\mathbf{r}_2^* = \mathbf{r}_2^*(r_1)$ is the value of \mathbf{r}_2 , given r_1 , for which $\phi(r_1, \mathbf{r}_2)$ attains its minimum. The approximation (12) is valid for r_1 in the domain defined by $1 + 2\beta_1 r_1 + 2\boldsymbol{\beta}'_2 \mathbf{r}_2^*(r_1) - \boldsymbol{\beta}' \boldsymbol{\beta} > 0$ and has a relative error of $O(T^{-1})$ in this domain.

To compute the marginal density approximant (12), we first evaluate \mathbf{r}_2^* for the required values of r_1 by direct optimization of $\phi(r_1, \mathbf{r}_2)$. This is a simple and inexpensive computation in practice. It is most useful to commence calculations for $r_1 = \beta_1$ at which point $\mathbf{r}_2^* = \bar{\mathbf{r}}_2 = \boldsymbol{\beta}_2$ by inspection of $\phi(r_1, \mathbf{r}_2)$. The same point can be used as a starting value in the iteration to minimize $\phi(r_1, \mathbf{r}_2)$ as we move away from $r_1 = \beta_1$, and successive optima can similarly be used in later iterations. By continuity, we should always be close to the minimum at the start of each iteration, provided the grid of r_1 values is sufficiently fine. No problems of multiple minima were encountered in this particular application of the algorithm. Once $\mathbf{r}_2^* = \mathbf{r}_2^*(r_1)$ is found, the function (12) itself can be directly computed.

Marginal densities based on (12) were computed to explore the form of these densities for variations in the number of endogenous variables in the structural equation and to examine the sensitivity of these densities to various changes in the parameters on which they depend. Some graphs that illustrate these effects are shown in Figures 1–7. Each of these graphs refers to an equation with $n = 3$ (that is, an equation with $n + 1 = 4$ endogenous variables) except those in Figure 5, which show the densities for values of n ranging from 2 to 7. The case of $n = 2$ has not been considered in any detail here because results for that case were obtained by numerical integration and reported in Phillips (1980b). Monte Carlo results for the same model with $n = 2$ (the three endogenous variable case) have been obtained by Richardson and Rohr (Chapter 2, this volume).

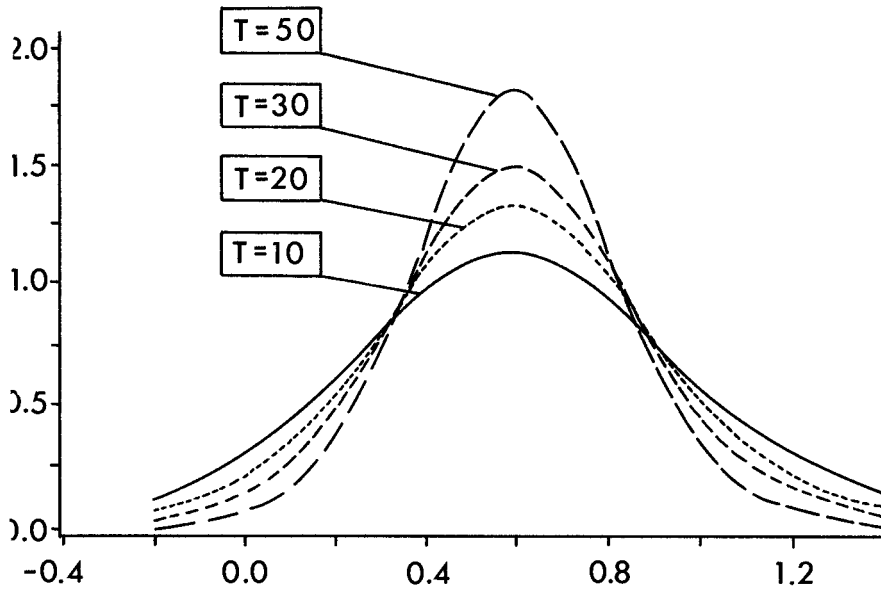
Some features that emerge from the densities graphed in Figures 1 to 7 are as follows:

1. For comparable parameter values, the marginal distribution appears to concentrate as $T \rightarrow \infty$ more slowly when $n = 3$ than when $n = 2$ or $n = 1$. This can be seen by comparing the apparent rate of convergence in Figure 1 with that of Figures 1 and 5 of Phillips (1980b) (the latter two figures are not reproduced here). The differences seem to be more marked between $n = 2$ and $n = 1$ (and $n = 3$ and $n = 1$) than between $n = 3$ and $n = 2$.

2. The density is particularly sensitive to the degree of correlation (ρ) in the matrix of products of reduced form coefficients, $\mathbf{M} = \bar{\boldsymbol{\Pi}}_{22}'\bar{\boldsymbol{\Pi}}_{22}$ (see Figure 2). This confirms similar results noted in Phillips (1980b). The dispersion of the density also increases with $|\rho|$. Since \mathbf{M} approaches singularity as $|\rho| \rightarrow 1$ and the equation becomes unidentifiable, this behavior accords with what we might expect from intuition. The central tendency of the distribution seems to be more sensitive to negative than positive values of ρ , a phenomena also noted in Phillips (1980b). The important factor in this phenomena is that the sensitivity occurs when

Figure 1. Densities of β_{IV} for various values of T when

$$n = 1, \beta_1 = 0.6, \beta_2 = 0.6, \beta_3 = 0.6, \mathbf{M} = \begin{bmatrix} 4.0 & 3.6 & 3.24 \\ 3.6 & 4.0 & 3.6 \\ 3.24 & 3.6 & 4.0 \end{bmatrix}, L = 3.$$

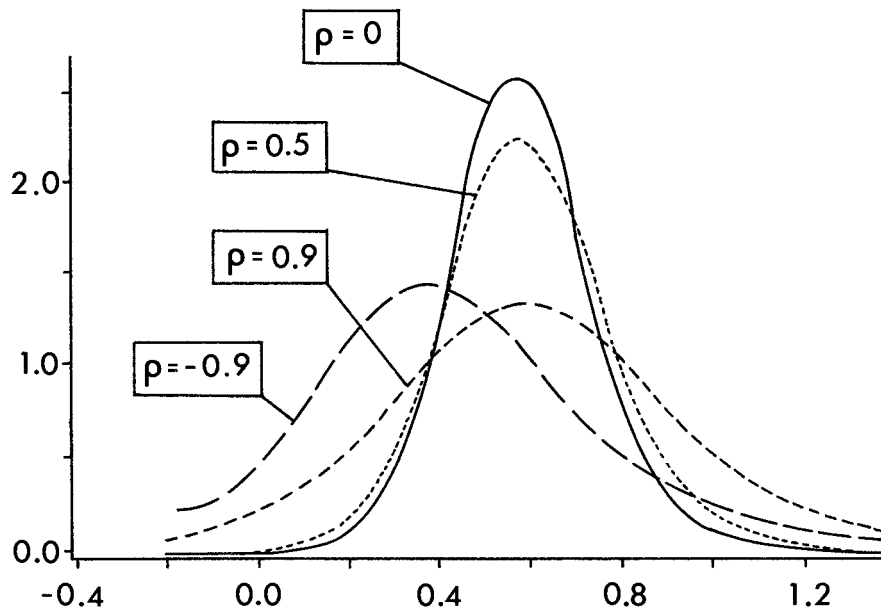


β_1 and ρ are of different signs. Thus, when the coefficients β_i and ρ all have the same sign, the common set of exogenous variables are compatible as instruments for Y_2 in the regression and the marginal distributions appear to be adequately centered; but when β_i and ρ take opposite signs, the exogenous variables are less compatible as instruments for the columns of Y_2 and the marginal distributions become less well centered about the true coefficients.

3. The dispersion and central tendency of the distribution can be very sensitive to the relative magnitude (μ) of the lengths of the reduced form coefficient vectors, as is clear from Figure 3. In other cases it is not, as in Figure 4. The rather dramatic difference between the two cases illustrated in Figures 3 (with μ varying) and 4 (with λ varying) can be directly explained from the analytic formula for the density given in (12). As λ increase in the matrix \mathbf{M} , it is clear that the parameter \bar{m}_{11} increases, and from the form of $\phi(r_1, r_2^*)$ in the exponential factor of (12) it follows that the density will display greater concentration about β_1 . This is essentially equivalent to a direct increase in the size of what would be the

Figure 2. Densities of β_{1IV} for various values of correlation ρ in M when

$$n = 3, \beta_1 = 0.6, \beta_2 = 0.6, \beta_3 = 0.6, \mathbf{M} = \begin{bmatrix} 4.0 & \rho 4.0 & 0 \\ \rho 4.0 & 4.0 & 0 \\ 0 & 0 & 4.0 \end{bmatrix}, L = 3, T = 20.$$

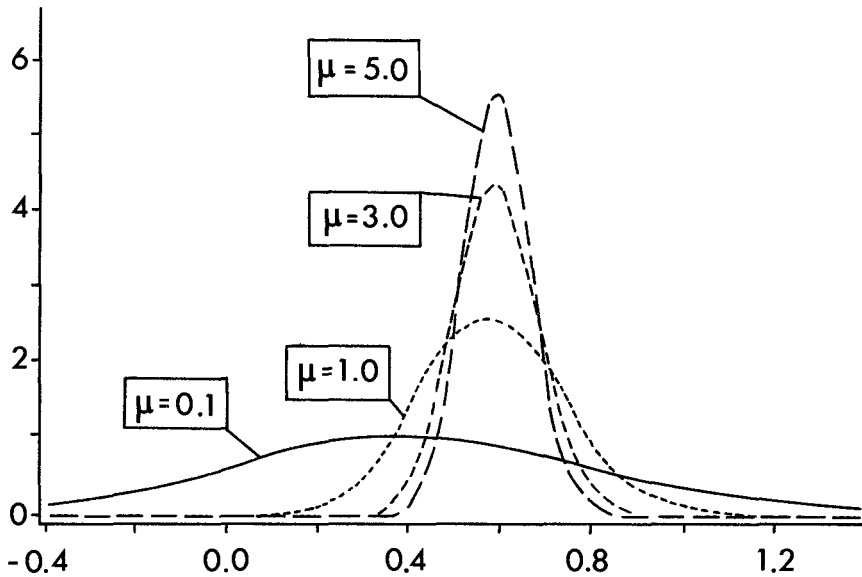


concentration parameter ($\mu^2 = T\bar{m}_{11}$) when $n = 1$ (the two endogenous variable case). As λ increases on the other hand, there is no such direct effect on the distribution of β_{1IV} . In fact, the graphs seem to suggest the contrary: that the increase in the concentration about β_3 in the distribution of β_{3IV} , which will result from the increase in λ (corresponding to the effect of an increase in μ on the distribution of β_{1IV} just discussed), may be achieved partially at the cost of a slight reduction in the concentration of the distribution of β_{1IV} (see, in particular, Figure 4). This is an effect that warrants some further investigation.

4. Figure 5 illustrates the effects of increasing the number of endogenous variables, *ceteris paribus*, on the marginal density of β_{1IV} . The effect is clear and monotonic in this case as a decrease in the precision of estimation. Further exploration of this case would also be of interest, for example, with the parameter values $\beta_1 = 0.6, \beta_i = 0$ ($i = 2, \dots, n$), which would

Figure 3. Densities of β_{1IV} for various values of μ in M when

$$n = 3, \beta_1 = 0.6, \beta_2 = 0.6, \beta_3 = 0.6, \mathbf{M} = \begin{bmatrix} \mu 4.0 & 0 & 0 \\ 0 & 4.0 & 0 \\ 0 & 0 & 4.0 \end{bmatrix}, L = 3, T = 20.$$

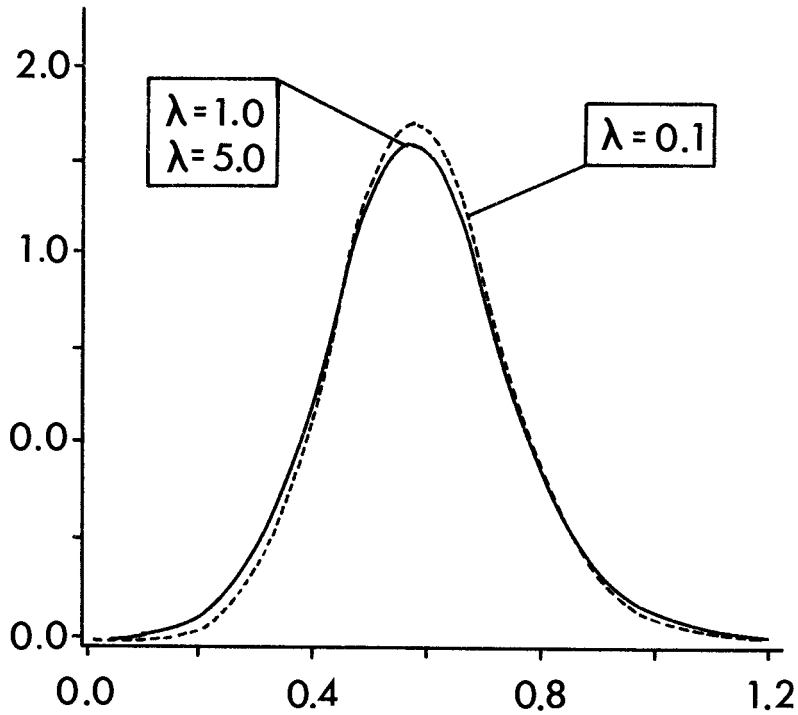


correspond to the erroneous inclusion of additional endogenous variables as regressors in the equation. Note that the recorded reduction in precision of estimation in this case accords with known results for the classical regression model.

5. The effects of variations in the coefficients of the other endogenous variables in the equation is explored in Figures 6 and 7. Once again we notice some rather interesting differences between the two cases. In both cases, as $\beta_2 - \beta_1$ increases and as $\beta_3 - \beta_1$ increases, the dispersion of the marginal distribution of β_{1IV} increases rapidly. However, as $\beta_2 - \beta_1$ increases, the bias becomes positive; when it decreases, the bias becomes negative. By contrast, there appear to be no major bias effects as $\beta_3 - \beta_1$ changes. The difference between the two cases arises from the difference between the correlation pattern in the matrix \mathbf{M} of cross products of reduced form coefficients. This seems to confirm the point made earlier in (2) and noted first in Phillips (1980b) that the correlation pattern of this matrix has a very important influence on the form of the marginal densities.

Figure 4. Densities of β_{1IV} for various values of λ in M when

$$n = 3, \beta_1 = 0.6, \beta_2 = 0.6, \beta_3 = 0.6, \mathbf{M} = \begin{bmatrix} 4.0 & 0 & 0 \\ 0 & 4.0 & 0 \\ 0 & 0 & \lambda 4.0 \end{bmatrix}, L = 3, T = 20.$$

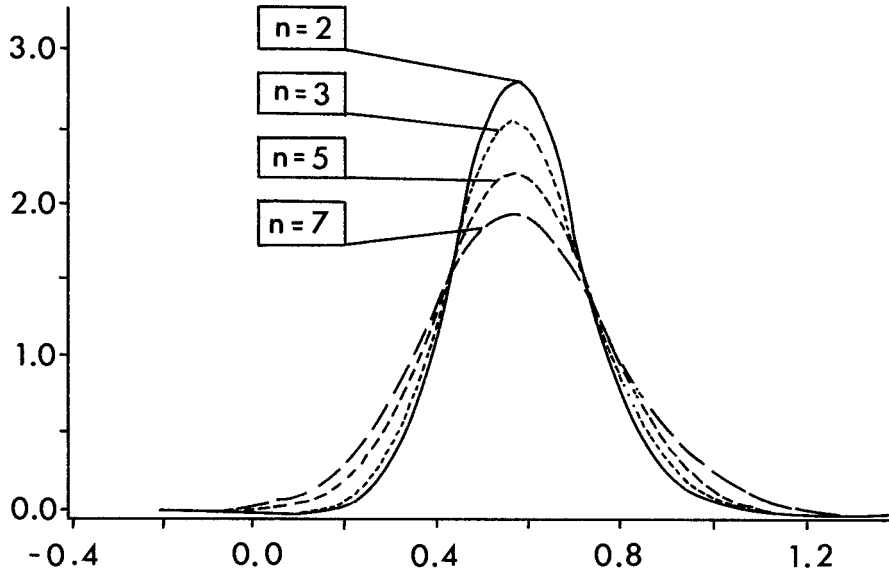


IV. EXACT DENSITIES IN SOME LEADING CASES

The exact joint density function given in (10) admits much simpler representations in certain leading cases. Some of these were already discussed in Section III of Phillips (1980b). Somewhat earlier, and before the general form of the exact density (10) was known, Basmann (1974) pointed out that the leading term in the multiple series could be obtained under a certain null hypothesis concerning the parameters. This built on the work by Basmann (1963) and Kabe (1964), which dealt with a specialized, leading three equation case.

In this section we shall give some further leading results for the general case of $n + 1$ equations.

Figure 5. Densities of β_{IV} for various values of n when $\beta_i = 0.6(i = 1, \dots, n)$, $\mathbf{M} = \text{diag}[4.0, \dots, 4.0]$, $L = 3$, $T = 20$.



CASE 1: $\beta = \mathbf{0}$

From (10) we deduce that

$$\text{pdf}(\mathbf{r}) = \frac{\text{etr}\{-(T/2)\bar{\mathbf{\Pi}}'_{22}\bar{\mathbf{\Pi}}_{22}\}\Gamma_n[(L+n+1)/2]}{\pi^{n/2}[\det(\mathbf{I} + \mathbf{r}\mathbf{r}')^{(L+n+1)/2}\Gamma_n[(L+n)/2]} \cdot {}_1F_1\{(L+n+1)/2, (L+n)/2; (T/2)\bar{\mathbf{\Pi}}_{22}(\mathbf{I} + \mathbf{r}\mathbf{r}')^{-1}\bar{\mathbf{\Pi}}'_{22}\} \quad (13)$$

When, in addition, $\bar{\mathbf{\Pi}}_{22} = \mathbf{0}$, this reduces to the leading term in the complete multiple series for the exact density [that is, the leading term of (10)], namely,

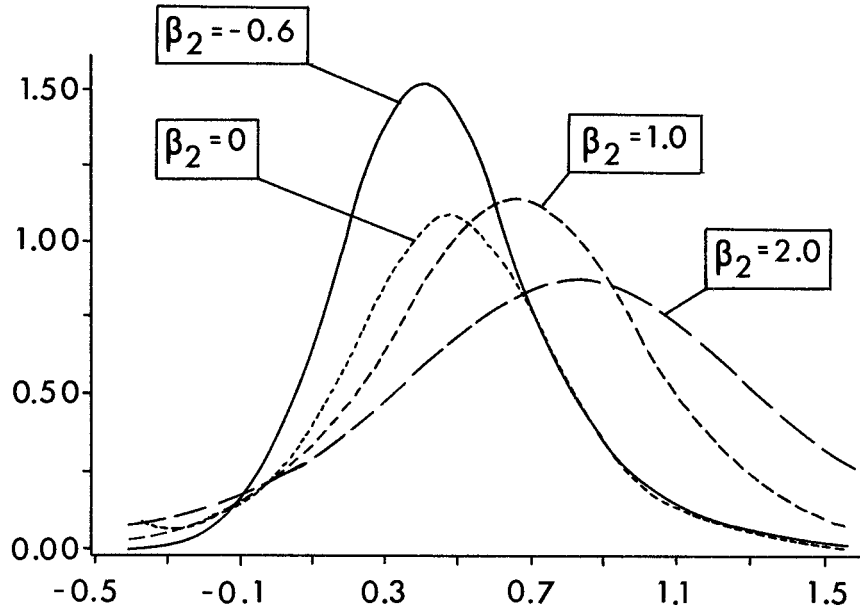
$$\begin{aligned} \text{pdf}(\mathbf{r}) &= \frac{\Gamma_n[(L+n+1)/2]/\Gamma_n[(L+n)/2]}{\pi^{n/2}(1 + \mathbf{r}'\mathbf{r})^{(L+n+1)/2}} \\ &= \frac{\Gamma[(L+n+1)/2]}{\pi^{n/2}\Gamma[(L+1)/2](1 + \mathbf{r}'\mathbf{r})^{(L+n+1)/2}} \end{aligned} \quad (14)$$

The probability density function (14) was given for the two-stage least squares case by Basmann (1974), who also deduced the leading marginal densities, which follow directly from (14) by a simple integration, namely,

$$\text{pdf}(r_i) = \frac{\Gamma[(L+2)/2]}{\pi^{1/2}\Gamma[(L+1)/2](1 + r_i^2)^{(L+2)/2}} \quad (15)$$

Figure 6. Densities of β_{1IV} for various values of β_2 when

$$n = 3, \beta_1 = 0.6, \beta_2 = 0.6, \mathbf{M} = \begin{bmatrix} 4.0 & 3.6 & 3.24 \\ 3.6 & 4.0 & 3.6 \\ 3.24 & 3.6 & 4.0 \end{bmatrix}, L = 3, T = 20.$$



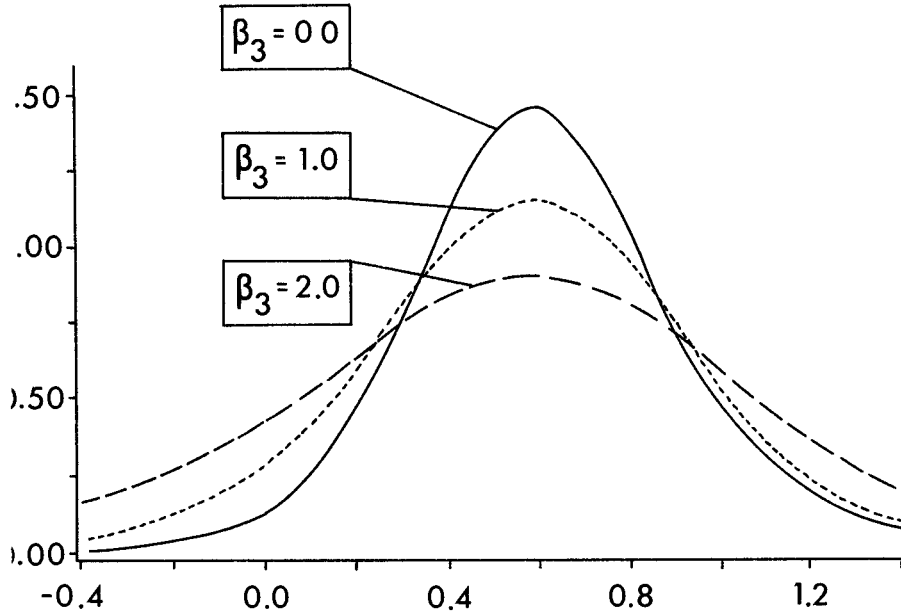
The leading marginal density function (15) has integer moments up to order $L = K_3 - n$ (the number of surplus instruments).

The matrix argument ${}_1F_1$ function in (13) is readily calculated for the three endogenous variable case ($n = 2$). Specifically, Herz (1955) and Muirhead (1975) give a series representation of the ${}_1F_1$ function in terms of the same function with a scalar argument. The form that (13) then takes is as follows:

$$\begin{aligned} \text{pdf}(r_1, \mathbf{r}_2) &= \frac{\text{etr}\{- (T/2) \bar{\Pi}'_{22} \bar{\Pi}_{22}\} \Gamma_2[(L+3)/2]}{\pi \Gamma_2[(L+2)/2] (1 + \mathbf{r}'\mathbf{r})^{(L+3)/2}} \\ &\cdot \sum_{j=0}^{\infty} \frac{[(L+3)/2]_j (1/2)_j}{[(L+1)/2]_j [(L+2)/2]_{2,j}!} \\ &\cdot {}_1F_1\{[(L+3)/2] + j, [(L+2)/2] + 2j; \\ &\quad T/2[\text{tr}(\bar{\Pi}'_{22} \bar{\Pi}_{22}) - (\mathbf{r}' \bar{\Pi}'_{22} \bar{\Pi}_{22} \mathbf{r}) / (1 + \mathbf{r}'\mathbf{r})]\} \\ &\cdot \{-(T/2)^2 [\det(\bar{\Pi}'_{22} \bar{\Pi}_{22})] / (1 + \mathbf{r}'\mathbf{r})\}^j \end{aligned} \quad (16)$$

Figure 7. Densities of β_{11v} for various values of β_3 when

$$n = 3, \beta_1 = 0.6, \beta_2 = 0.6, \mathbf{M} = \begin{bmatrix} 4.0 & 3.6 & 3.24 \\ 3.6 & 4.0 & 3.6 \\ 3.24 & 3.6 & 4.0 \end{bmatrix}, L = 3, T = 20.$$



The marginal densities can now be extracted from (16) by integration. Since the derivations are lengthy, I only give the final formulae here. Specifically we obtain

$$\begin{aligned} \text{pdf}(r_1) &= \frac{[(L+1)/2] \exp\{-T/2(m_{11} + m_{22})\}}{\pi(1+r_1^2)^{(L+2)/2}} \\ &\cdot \sum_{j,k=0}^{\infty} \sum_{u+2v+w=k} \frac{[(L+3)/2]_j (1/2)_j [(L+3)/2+j]_k}{[(L+1)/2]_j [(L+2)/2]_{2j} [(L+2)/2+2j]_k} \\ &\cdot \frac{\Gamma[(L+2)/2+j+k-v-w] \Gamma[v+w+(1/2)] [-(T/2)^2(m_{11}m_{22} - m_{12}^2)]^j}{\Gamma[(L+3)/2+j+k] j! u!(2v)! w!} \\ &\cdot \frac{[(T/2)\{m_{11} + m_{22}(1+r_1^2)\}]^v (Tr_1 m_{12})^{2v} [(T/2)m_{11}]^w}{(1+r_1^2)^{j+k-u-w}} \end{aligned} \quad (17)$$

where we use the notation $\mathbf{M} = (m_{ij}) = \bar{\Pi}'_{22} \bar{\Pi}_{22}$. This result generalizes all previous known formula for exact marginal densities in leading simultaneous equations cases. The following sections present special cases of (17) already in the literature.

CASE 2. $\beta = \mathbf{0}$, $L = 1$, $m_{12} = 0$

This function was presented by Kabe (1964), equation (4.14), where the multiple series is in a somewhat different form).

$$\begin{aligned} \text{pdf}(r_1) &= \frac{\exp\{-(T/2)(m_{11} + m_{22})\}}{\pi(1 + r_1^2)^{3/2}} \\ &\cdot \sum_{j,k=0}^{\infty} \sum_{u+w=k} \frac{(j+1)(1/2)_j(j+2)_k \Gamma[j+k-w+(3/2)] \Gamma[w+(1/2)]}{j!(3/2)_j [2j+(3/2)]_k \Gamma[j+k+2] u! w!} \\ &\cdot \frac{(-1)^j [(T/2)m_{11}]^{j+w} [(T/2)m_{22}]^j [(T/2)\{m_{11} + m_{22}(1 + r_1^2)\}]^u}{(1 + r_1^2)^{j+k-u-w}} \end{aligned}$$

CASE 3. $\beta = \mathbf{0}$, $L = 1$, $m_{22} = 0$, $m_{12} = 0$.

This function was presented by Basmann [1963, equation (4.18)].

$$\text{pdf}(r_1) = \frac{\exp(-(T/2)m_{11})}{2(1 + r_1^2)^{3/2}} \sum_{k=0}^{\infty} \frac{1}{k!} {}_1F_1(1/2, 3/2 + k, (T/2)m_{11}) \left[\frac{Tm_{11}}{2(1 + r_1^2)} \right]^k$$

CASE 4. $\beta = \mathbf{0}$, $L = 1$, $m_{11} = 0$, $m_{22} = 0$, $m_{12} = 0$.

This function was presented by Basmann [1963, equation (4.13)].

$$\text{pdf}(r_1) = \frac{1}{2(1 + r_1^2)^{3/2}}$$

The general expression (17) for the marginal density may be used for numerical computations or, alternatively, the joint density (16) can be summed and the marginal densities extracted by a one dimensional quadrature. At present the author has no numerical experience to report with either of these approaches. However, the formulae suggest that numerical computations of the exact marginal densities are now possible for these leading cases when $n = 2$.

Working from the general leading case (Case 1), further analytic results can be obtained. We illustrate with the following final example without going into all the algebra.

CASE 5. $\beta = \mathbf{0}$, $\bar{\Pi}_{22} = \text{diag}(0, \dots, \Pi_{aa}, \dots, 0)$.

We let r_a be the a th element of \mathbf{r} and use \mathbf{r}_b to denote the vector of the remaining elements. Π_{aa}^2 is a concentration coefficient representing the squared length of the coefficients of \mathbf{Z}_2 in the reduced form equation for the a th endogenous variable. $T\Pi_{aa}^2$ therefore corresponds to the usual concentration parameter in the two endogenous variable case (when $n = 1$). Then, from (13)

$$\begin{aligned}
\text{pdf}(r_a, \mathbf{r}_b) &= \frac{\exp[-(T/2)\mathbf{\Pi}_{aa}^2]\Gamma_n[(L+n+1)/2]}{\pi^{n/2}(1+\mathbf{r}'_b\mathbf{r}_b+r_a^2)^{(L+n+1)/2}\Gamma_n[(L+n)/2]} \\
&\quad \cdot {}_1F_1\left[(L+n+1)/2, (L+n)/2; \frac{T\mathbf{\Pi}_{aa}^2(1+\mathbf{r}'_b\mathbf{r}_b)}{2(1+\mathbf{r}'_b\mathbf{r}_b+r_a^2)}\right] \\
&= \frac{\exp[-(T/2)\mathbf{\Pi}_{aa}^2]\Gamma_n[(L+n+1)/2]}{\pi^{n/2}\Gamma_n[(L+n)/2]} \sum_{k=0}^{\infty} \frac{[(L+n+1)/2]_k}{[(L+n)/2]_k k!} \\
&\quad \cdot \left[\frac{T\mathbf{\Pi}_{aa}^2}{2}\right]^k \cdot \frac{(1+\mathbf{r}'_b\mathbf{r}_b)^k}{(1+\mathbf{r}'_b\mathbf{r}_b+r_a^2)^{k+(L+n+1)/2}}
\end{aligned}$$

We transform using $\mathbf{r}_b = (1+r_a^2)^{1/2}\mathbf{z}$ where \mathbf{z} has dimension $n-1$. The Jacobian is $(1+r_a^2)^{(n-1)/2}$ and the marginal density is given by

$$\begin{aligned}
\text{pdf}(r_a) &= \frac{\exp[-(T/2)\mathbf{\Pi}_{aa}^2]\Gamma_n[(L+n+1)/2]}{\mathbf{\Pi}^{n/2}\Gamma_n[(L+n)/2](1+r_a^2)^{(L+2)/2}} \\
&\quad \cdot \sum_{k=0}^{\infty} \frac{[(L+n+1)/2]_k}{[(L+n)/2]_k k!} \left(\frac{T\mathbf{\Pi}_{aa}^2}{2}\right)^k \\
&\quad \sum_{l=0}^k \binom{k}{l} \frac{1}{(1+r_a^2)^l} \int \frac{(\mathbf{z}'\mathbf{z})^{k-l} d\mathbf{z}}{(1+\mathbf{z}'\mathbf{z})^{k+(L+n+1)/2}}
\end{aligned} \tag{18}$$

The integral in the final expression can be reduced analytically in the usual way by writing $\mathbf{z}' = (\mathbf{z}_c, z_d)$ where z_d is scalar and by changing the variable to x via the transformation $z_d = (1+\mathbf{z}'_c\mathbf{z}_c)^{1/2}\tan x$. We can proceed in this way, taking each component of \mathbf{z} separately, until the integral is evaluated.

V. FINAL COMMENTS

The method discussed in Section II should have fairly general applicability to the problem of extracting marginal densities and should be useful in the wider context of characterizing multidimensional distributions in both the sampling theoretic and Bayesian approaches. The numerical illustrations of the method as it is applied to the general single equation IV estimator in Section III show that the technique works well even for high dimensional cases and is successful in isolating the parameters that are most critical in determining the form of the marginal densities.

The exact density results for leading cases in Section IV extend the earlier work of Basmann and Kabe. The results are also amenable to computation, at least in the three endogenous variable case. Further work, which is currently in progress, will analyze the effects of misspecification within the same general single equation set up.

POSTSCRIPT

Since this article was written the author has developed some technical machinery which facilitates the analytic determination of marginal densities in full generality. Details of this research will be reported later.

ACKNOWLEDGMENTS

My thanks to Glenna Ames for her skill and time in typing the manuscript. The research reported in this article was supported by the National Science Foundation under grant No. SES 8007571. The graphs in Section III were prepared and related programming work was conducted by Sidnie Feit, for whose help I am most grateful.

NOTES

1. It seems likely that a new algorithm for computation will soon become available based on the work of Towber (1979) on the irreducible polynomial representations of the general linear group.
2. Careful analysis of this problem has shown that series rearrangements prior to summation allow adequate convergence to be achieved using fewer terms in this case.
3. Of course, even in the two endogenous variable case, the usual analysis of case (1) starts off in a multidimensional framework. For example, Basmann (1961) initiates the derivations that lead ultimately to the density of the scalar coefficient estimator by considering the multivariate normal density of the least squares estimators of the reduced form coefficients. Bergstrom (1962) commenced his derivations with the joint density of the sample data vector in T dimensions. Kabe (1963) derived the same exact density as Basmann (1961) by working with the noncentral Wishart representation of the second moments of the data. His work formed the basis for the general two equation exact results obtained later by Richardson (1968) and Sawa (1969).
4. Kloek and Van Dijk (1978) do present some Monte Carlo estimates of marginal densities in their model in addition to moment estimates.
5. In certain cases, it will be more convenient to represent the density as a multiple integral of the Fourier type with an exponential kernel of the form $\exp\{i\lambda(T)\phi(\mathbf{r}, \mathbf{b})\}$. In such cases, the major contribution to the value of the integral as $T \rightarrow \infty$ can come from points in the domain where the smoothness conditions on ϕ and g fail as well as stationary points of ϕ . An introduction to the asymptotic treatment of multiple integrals of this type is given by Bleistein and Handelsman (1976) in 8.4.

REFERENCES

- Basmann, R. L. (1961) A note on the exact finite sample frequency functions of generalized classical linear estimators in two leading overidentified cases, *JASA* 56, 619-636
- Basmann, R. L. (1963) A note on the exact finite sample frequency functions of generalized classical linear estimators in a leading three equation case. *JASA* 58, 161-171

- Basmann, R. L. (1974). Exact finite sample distributions and test statistics: a survey and appraisal. In M. D. Intriligator and D. A. Kendrick (eds.), *Frontiers of Quantitative Economics*, Vol. II, Ch. 4, Amsterdam. North-Holland.
- Bergstrom, A. R. (1962). The exact sampling distributions at least squares and maximum likelihood estimators of the marginal propensity to consume, *Econometrica* 30, 480–490.
- Bleistein, N. and R. A. Handelsman (1976). *Asymptotic Expansions of Integrals*, New York: Holt, Rinehart & Winston.
- Constantine, A. G. and R. J. Muirhead (1976). Asymptotic expansions for distributions of latent roots in multivariate analysis, *J. Multivar. Anal.* 6, 369–391.
- Chikuse, Y. (1981). Distributions of some matrix variates and latent roots in multivariate Behrens-Fisher discriminant analysis, *Ann. Statist.* 9, 401–407.
- Davis, A. W. (1979). Invariant polynomials with two matrix arguments extending the zonal polynomials. applications to multivariate distribution theory, *Ann. Inst. Stat. Math.*, 31, A, 465–485.
- Davis, A. W. (1980). Invariant polynomials with two matrix arguments extending the zonal polynomials. In P. R. Krishnaiah (ed.), *Multivariate Analysis*, Amsterdam. North-Holland.
- Herz, C. S. (1955). Bessel functions of matrix argument, *Ann. Math.* 61, 474–523.
- Holly, A. and P. C. B. Phillips (1979). A saddlepoint approximation to the distribution of the k -class estimator in a coefficient in a simultaneous system, *Econometrica* 47, 1527–1547.
- Hurwicz, L. (1950). Least squares bias in time series, In T. C. Koopmans (ed), *Statistical Inference in Dynamic Economic Models*, New York: Wiley.
- James, A. T. (1961). Zonal polynomials of the real positive definite symmetric matrices, *Ann. Math.* 74, 456–469.
- James, A. T. (1968). Calculation of zonal polynomial coefficients by use of the Laplace-Beltrami operator, *Ann. Math. Statist.* 39, 1711–1718.
- Kabe, D. G. (1963). A note on the exact distributions of the GCL estimators in two leading overidentified cases, *JASA* 58, 535–537.
- Kabe, D. G. (1964). On the exact distributions of the GCL estimators in a leading three equation case, *JASA* 59, 881–894.
- Kloek, T. and H. K. Van Dijk (1978). Bayesian estimates of equation system parameters. An application of integration by Monte Carlo, *Econometrica* 46, 1–19.
- McLaren, M. L. (1976). Coefficients of the zonal polynomials, *App. Stat.* 25, 82–87.
- Muirhead, R. J. (1975). Expressions for some hypergeometric functions of matrix argument with applications, *J. Multivar. Anal.* 5, 283–293.
- Muirhead, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results, *Ann. Stat.* 6, 5–33.
- Nagel, P. J. A. (1981). Programs for the evaluation of zonal polynomials, *Am. Stat.* 35, 53.
- Phillips, P. C. B. (1977). A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators, *Econometrica* 45, 1517–1534.
- Phillips, P. C. B. (1980a). Finite sample theory and the distributions of alternative estimators of the marginal propensity to consume, *Rev. Econ. Studies* 47, 183–224.
- Phillips, P. C. B. (1980b). The exact finite sample density of instrumental variable estimators in an equation with $n + 1$ endogenous variables, *Econometrica* 48, 861–878.
- Phillips, P. C. B. (1982). Best uniform and modified Padé approximants to probability densities in econometrics, presented to World Congress of Econometric Society, Aix-

- en-Provence, 1980 and published in W. Hildenbrand (Ed) "*Advances in Econometrics*," Cambridge University Press, pp 123–167
- Rhodes, G. F. (1981). Exact density functions and approximate critical regions for likelihood ratio identifiability test statistics, *Econometrica*, 49, pp. 1035–1056.
- Richardson, D. H. (1968). The exact distribution of a structural coefficient estimator, *JASA* 63, 1214–1226.
- Richardson, D. H. and R. J. Rohr (1983) "An experimental study of two stage least squares distributions in equations containing three endogenous variables" In R. Basman and George F. Rhodes, Jr. (eds), *Advances in Econometrics*, vol. 2, Greenwich, Ct.: JAI Press, pp. 25–50.
- Sargan, J. D. (1976). Econometric estimators and the Edgeworth approximation, *Econometrica* 44, 421–428, and Erratum, *Econometrica* 15, 272.
- Sawa, T. (1969). The exact finite sampling distribution of ordinary least squares and two stage least squares estimator, *JASA* 64, 923–936.
- Slater, L. J. (1965). Confluent hypergeometric functions, In M. Abramowitz and I. A. Stegun (eds), *Handbook of Mathematical Functions*, New York: Dover.
- Towber, J. (1979). Young symmetry, the flag manifold and representations of $GL(n)$, *J. Algebra* 61, 414–462.