

## 10

Operational Algebra and Regression  $t$ -Tests

Peter C. B. Phillips

## Introduction

In an article published by *Econometrica*, Rex Bergstrom (1962) derived the exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume in a two-equation Keynesian model of income determination. This article emphasized the importance of a mathematical study of the small-sample behavior of econometric estimators. Bergstrom's objective was to evaluate the performance characteristics of the new simultaneous equations estimators in relation to least squares and his conclusion, at least for the model studied, was unambiguous: "... for samples of ten or more observations ... the maximum likelihood estimator of the marginal propensity to consume is the 'better' general purpose estimator of this parameter."

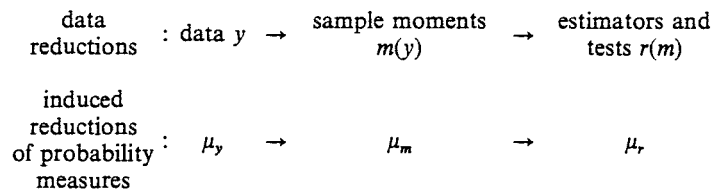
The Bergstrom article, together with a related paper by Basmann (1961), created a field of research that became known as finite-sample theory in econometrics. Progress in the field was intermittent through the 1960s and 1970s, with few spectacular advances. This is in part explained by the mathematical difficulty of the research, but another factor was the rapid development in computer technology during this period. This opened the way to large-scale simulation exercises that provided quick and easy numerical information about the sampling characteristics of estimators and tests in stochastic environments similar to those in which they were to be used. The 1980s witnessed even greater enhancements in computer technology. With the advent of cheap personal computers and workstations capable of 32-bit arithmetic and with simulation-based statistical methods like the bootstrap, it seems likely that attention will continue to move away from mathematical studies of the Bergstrom type.

Nevertheless, mathematical studies remain fascinating and have continued to attract a few dedicated researchers. Perhaps the biggest obstacle to research in the field is the specificity of individual studies. Each new estimator or test seems to bring with it a new set of mathematical difficulties that sustained effort, ingenuity, and technical skill are not always sufficient to resolve. As a result, graduate students and young researchers are naturally more easily drawn into asymptotic analyses where there is the rich reservoir of theorems and methods from probability theory to draw upon when there is a need to demonstrate hard quantitative conclusions, and into simulation exercises when there is a need to illustrate sampling performance. The

latter exercises have themselves been facilitated by the widespread availability of matrix programming languages such as GAUSS (1989), which is one of the more popular packages among econometricians. Software packages like GAUSS allow formulas to be programmed much as they are written down in matrix format, so that the transition from econometric formula to computer program is enormously simplified. Moreover, the graphics facilities that GAUSS supports are sophisticated and easy to use. With such software, it is nowadays usual practice to have simulations up and running with accompanying graphics within a few days of developing the theory or, indeed, as work on the theory is itself under way. What might have taken six months developmental work in the 1960s is now done in a week. And the programming skills that support the use of software like GAUSS are now as much a part of the econometrician's tool kit as linear algebra.

The model that I have just described of theoretical quantitative research underpinned by asymptotics and simulation has become widespread in recent years. It might easily be taken as the standard research paradigm of today's young econometric theorist. It seems likely that it will become almost universal during the 1990s.

This combination of asymptotics and simulation is usually highly productive and informative, but it is not without drawbacks, perhaps the most important of which is that it diverts attention from some vital issues. Econometrics, like statistics, is concerned with data reduction. Estimators and test statistics are summary measures and the formulas that give rise to them carry with them certain characteristics. The characteristics of the mapping from data to statistics are often of primary importance in understanding the statistical properties of a given procedure. Indeed, the physical form of the statistic as a function of the data itself induces an operation on the probability law of the sample from which the statistic is derived. The situation is exhibited in the following scheme:



In many cases, the probability measure  $\mu_y$  of the data will be absolutely continuous with respect to Lebesgue measure and the transition at the second level can be formulated in terms of probability densities as PDF ( $y$ ) → PDF ( $m$ ) → PDF ( $r$ ). The fundamental question addressed in a mathematical study that seeks to derive PDF ( $r$ ) is this: can the operations on PDF ( $y$ ) that are induced by the data maps  $y \rightarrow m \rightarrow r$  be formalized algebraically and reduced to give a solution in a closed form or a series representation?

Of course, understanding the transition  $y \rightarrow m \rightarrow r$  and what this reduction loses in the way of information in the sample is critical to virtually all statistical theory. But solutions are available only in specific instances and no one seems yet to have attempted a general theory. The present paper is designed to offer some thoughts on how to tackle this general problem and to illustrate an algebraic approach that seems promising. Let it be clear at the outset that I am not proposing an operational

procedure for practical applications. So what I have to say will in no way be competitive with the ongoing research paradigm of “asymptotic theory plus illustrative simulations” that I described above. However, the idea that I am going to put forward has the merit of providing a general mathematical framework for solving distributional problems. In this sense, it follows the spirit of the original Bergstrom (1962) study and eschews the research direction that has been followed by much of the profession since then.

On matters of notation I shall use the symbol “ $\equiv$ ” to signify equivalence in distribution, and “ $\sim$ ” to signify asymptotic equivalence.  $Lt(\cdot)$  and  $PDF(\cdot)$  represent the Laplace transform and probability density of a distribution,  $\Gamma(\cdot)$  is the gamma function and  $B(\cdot, \cdot)$  the beta function.  $\chi_n^2(\cdot)$  is the density of the chi-squared distribution with  $n$  degrees of freedom and  $N(0, 1)$  is the standard normal density, usually with argument  $r$ , viz.  $(2\pi)^{-1/2} e^{-r^2/2}$ .

### Functions of Differential Operators, Pseudodifferential and Fourier Integral Operators

At the outset it is useful to start with a space of functions that have nice properties, not all of which are really needed. Accordingly, we let  $C_\infty$  be the space of infinitely differentiable functions  $f(x)$  on the real line  $(-\infty, \infty)$ . If  $D = d/dx$  represents the usual operation of differentiation then  $D$  may be interpreted as the mapping  $D: C_\infty \rightarrow C_\infty$  defined by  $Df(x) = f'(x) \in C_\infty, \forall f \in C_\infty$ . In a similar way we may attach a meaning to polynomial functions of  $D$  such as  $p_n(D): C_\infty \rightarrow C_\infty, p_n(D) = \sum_{i=0}^n a_i D^i$  which are defined according to the relation  $p_n(D)f(x) = \sum_{i=0}^n a_i f^{(i)}(x)$  setting  $f^{(0)} = f$  and using  $f^{(i)}$  to signify the  $i$ th derivative of  $f$ . Such a class of operators is useful in solving ordinary differential equations with constant coefficients.

It is also useful to attach a meaning to nonintegral powers of  $D$ . This leads to the concept of fractional differentiation (and integration). The simplest approach is to rely on the gamma integral for a complex power which yields the formula

$$D^{-\alpha}f(x) = \Gamma(\alpha)^{-1} \int_0^\infty [e^{-Dt}f(x)]t^{\alpha-1} dt = \Gamma(\alpha)^{-1} \int_0^\infty f(x-t)t^{\alpha-1} dt, \quad (1)$$

which is valid for complex  $\alpha$  with  $\text{Re}(\alpha) > 0$  provided the integral converges. Note that the final integral representation in (1) is induced via the Taylor formula  $\exp(-Dt)f(x) = f(x-t)$ , which holds for  $f(\cdot)$  analytic; but this final representation takes on an independent life as a definition of  $D^{-\alpha}f(x)$  provided the integral converges. Moreover, quite general complex powers such as  $D^\mu$  may now be defined by setting  $D^\mu f(x) = D^{-\alpha}\{D^m f(x)\}$  with  $\mu = m - \alpha, \text{Re}(\alpha) > 0$  and  $m$  a non-negative integer. Fractional operators defined this way form a Weyl calculus (see for example Miller (1974)) and can be easily extended to matrices using the multivariate gamma integral in place of (1). Such extensions were introduced independently by the author (1985) and by Richards (1984) in quite different contexts. The use of such operators in resolving problems of distribution theory is illustrated in the author's papers (1984, 1985, 1986, 1987).

More general functions than powers of  $D$  can be accommodated through use of Fourier transforms. Suppose  $g(z)$  has Fourier transform  $\tilde{g}(p) = \int_{\mathbf{R}} e^{ipz}g(z) dz$ . Then we define the operator  $g(D)$  by its action on  $f$  through the inverse transform as

follows:

$$g(D)f(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} [e^{-ipD}f(x)]\tilde{g}(p) dp = (2\pi)^{-1} \int_{-\infty}^{\infty} f(x - ip)\tilde{g}(p) dp, \quad (2)$$

provided the integral converges.

It is easy to see that with this definition  $g(D) e^{ax} = e^{ax}g(a)$ , as in elementary calculus. Moreover

$$g(D) e^{ax}f(x) = e^{ax}g(D + a)f(x), \quad (3)$$

just as when  $g(D)$  is a polynomial. These rules for operating on exponentials can then be employed to deal with other elementary functions. For instance, if  $\text{Re}(\beta) > 0$  we have

$$f(D)(1 - x)^{-\beta} = (1/\Gamma(\beta)) \int_0^{\infty} \exp[-(1 - x)t] t^{\beta-1} f(t) dt, \quad (4)$$

which for  $f(D) = D^{\mu}$  leads to

$$D^{\mu}(1 - x)^{-\beta} = [\Gamma(\beta + \mu)/\Gamma(\beta)](1 - x)^{-\beta-\mu} = (\beta)_{\mu}(1 - x)^{-\beta-\mu}, \quad (5)$$

where  $(\beta)_{\mu} = \beta(\beta + 1) \dots (\beta + \mu - 1)$  is the forward factorial, with the convention that  $(0)_0 = 1$ . Note that the final formula then holds by analytic continuation irrespective of the values of  $\beta$  and  $\mu$ .

Formula (2) has a natural multivariate extension. Suppose  $x \in \mathbb{R}^n$  and let  $\partial x = \partial/\partial x$ . We define

$$g(\partial x)f(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} [e^{-ip' \partial x}f(x)]\tilde{g}(p) dp = (2\pi)^{-n} \int_{\mathbb{R}^n} f(x - ip)\tilde{g}(p) dp, \quad (6)$$

where  $g$  and  $\tilde{g}$  are again a Fourier transform pair. Since the Fourier transform of a generalized function always exists, it is often useful to work with generalized functions or ordinary functions that are used as generalized functions in this representation. For example, suppose in (6) that  $g(z) = \exp(b'z)$  and  $z$  is an  $n$ -vector. Then  $\tilde{g}(p) = (2\pi)^n \delta(p - ib)$  where  $\delta(\cdot)$  is the dirac generalized function (cf. Gelfand and Shilov (1964), pp. 169–90). We deduce from (6) that

$$g(\partial x)f(x) = \int_{\mathbb{R}^n} f(x - ip)\delta(p - ib) dp. \quad (7)$$

Upon deforming the contour of integration from  $\mathbb{R}^n$  to  $ib + \mathbb{R}^n$  in  $\mathbb{C}^n$ , we obtain

$$g(\partial x)f(x) = \int_{ib + \mathbb{R}^n} f(x - ip)\delta(p - ib) dp = f(x - i^2b) = f(x + b), \quad (8)$$

which corresponds with the usual Taylor formula, viz.  $e^{b' \partial x}f(x) = f(x + b)$ . The same

result can be obtained by using the Fourier transform  $\tilde{f}(p)$  of  $f(\cdot)$  rather than  $g(\cdot)$  in the definition (6). We would then have

$$g(\partial x)f(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} g(\partial x) e^{-ip'x} \tilde{f}(p) dp = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-ip'x} g(-ip) \tilde{f}(p) dp. \quad (9)$$

When  $g(z) = \exp(b'z)$  the right-hand side of (9) is trivially  $f(x + b)$ . Note finally that we can invert the Fourier transform in (9), giving us the representation

$$g(\partial x)f(x) = (2\pi)^{-n} \int_{\mathbb{R}^{2n}} e^{-ip'(x-\zeta)} g(-ip) f(\zeta) d\zeta dp. \quad (10)$$

As given above, (2), (6), and especially (9) are all closely related to the concept of a pseudodifferential operator. Suppose  $f$  has Fourier transform  $\tilde{f}$  in  $\mathbb{R}^n$ . Then a pseudodifferential operator  $k(x, i \partial x)$  is usually defined as

$$k(x, i \partial x)f(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} k(x, p) \tilde{f}(p) e^{-ix'p} dp. \quad (11)$$

Again the representation retains its meaning for generalized functions and their transforms. The operator  $k(x, i \partial x)$  is generally noncommutative in its arguments and the action of the operators  $x$  and  $\partial x$  is taken in the given order, i.e.  $k(x, i \partial x)$ , with the elements of  $\partial x$  acting before those of  $x$ . Note that in contrast to (6) we are now employing the Fourier transform of the operand  $f$  in (11) as distinct from that of the operator as in (6). Thus, (11) may be interpreted as a natural extension of (9).

There is now a vast literature on the properties of such operators and they are extensively used in the theory of partial differential equations. Treves (1982, volume 1) and Dieudonné (1988) provide a detailed study. An example of the use of a special case of (11) is the Laplace equation

$$\nabla^2 y(x) = f(x), \quad \nabla^2 = \partial x' \partial x, \quad (12)$$

which is solved by

$$y(x) = \nabla^{-2} f(x) = -(2\pi)^{-n} \int_{\mathbb{R}^n} e^{-ix'p} (p'p)^{-1} \tilde{f}(p) dp. \quad (13)$$

Partial differential equations with variable coefficients then lead to solutions that involve noncommutative operators of the form given in (9).

Further extensions of operators based on (9) are available in which  $x'p$  in the exponent is replaced by a real-valued  $C_\infty$  function  $\varphi(x, p)$ . Such extended operators are known as Fourier integral operators and they are discussed in Treves (1982, volume 2). We shall have no need for them in the present work.

We end this section with the following lemma that reports two useful results on nonlinear functions of operators induced by the standard normal density. We employ

the notation

$$N(0, 1/\hat{\partial}y) = (2\pi)^{-1/2}(\hat{\partial}y)^{1/2} \exp[-\frac{1}{2}r^2 \hat{\partial}y]. \quad (14)$$

**Lemma**

$$N(0, 1/\hat{\partial}y) e^{xy^m} = N(0, 1) e^y \sum_{j=0}^m \binom{m}{j} (-1)^j (-\frac{1}{2})_j (y - r^2/2)^{m-j}; \quad (15)$$

$$[N(0, n/\hat{\partial}x' \hat{\partial}x) e^{x'x/2}]_{x=0} = [N(0, n/\hat{\partial}y)(1 - 2y)^{-n/2}]_{y=0}; \quad (16)$$

where  $y$  is a scalar in (15) and  $x$  is an  $n$ -vector in (16).

**Proof.** To prove (15) we introduce the auxiliary variable  $\alpha$  and write

$$\begin{aligned} N(0, 1/\hat{\partial}y) e^{xy^m} &= [N(0, 1/\hat{\partial}y) \hat{\partial}\alpha^m e^{xy}]_{\alpha=1} \\ &= [(\hat{\partial}\alpha)^m N(0, 1/\hat{\partial}y) e^{xy}]_{\alpha=1} = [\hat{\partial}\alpha^m e^{xy} N(0, 1/\alpha)]_{\alpha=1} \\ &= [(\hat{\partial}\alpha)^m \{e^{xy} (2\pi)^{-1/2} \alpha^{1/2} e^{-\alpha^2/2}\}]_{\alpha=1} \\ &= (2\pi)^{-1/2} [\hat{\partial}\alpha^m (\alpha^{1/2} e^{x(y-r^2/2)})]_{\alpha=1} \\ &= (2\pi)^{-1/2} \left[ \sum_{j=0}^m \binom{m}{j} (\hat{\partial}\alpha^j \alpha^{1/2}) (\hat{\partial}\alpha^{m-j} e^{x(y-r^2/2)}) \right]_{\alpha=1} \\ &= (2\pi)^{-1/2} \left[ \sum_{j=0}^m \binom{m}{j} (-1)^j (-\frac{1}{2})_j \alpha^{1/2-j} (y - r^2/2)^{m-j} e^{x(y-r^2/2)} \right]_{\alpha=1} \\ &= (2\pi)^{-1/2} e^{-r^2/2} e^y \sum_{j=0}^m \binom{m}{j} (-1)^j (-\frac{1}{2})_j (y - r^2/2)^{m-j} \end{aligned} \quad (17)$$

as stated. To prove (16) we let  $\xi \equiv N(0, I)$  so that  $E(e^{x'\xi}) = e^{x'x/2}$  is the Laplace transform. Then

$$\begin{aligned} N(0, n/\hat{\partial}x' \hat{\partial}x) e^{x'x/2} &= N(0, n/\hat{\partial}x' \hat{\partial}x) E(e^{x'\xi}) \\ &= E[N(0, n/\hat{\partial}x' \hat{\partial}x) e^{x'\xi}] \\ &= E[N(0, n/\xi'\xi) e^{x'\xi}]. \end{aligned} \quad (18)$$

Next write  $\xi$  in polar coordinates as  $\xi = r^{1/2} \mathbf{h}$  where  $r = \xi'\xi \equiv \chi_n^2$  and  $\mathbf{h}$  is uniformly distributed on the unit sphere in  $\mathbb{R}^n$ . Then, upon evaluation at  $x = 0$ , (18) becomes

$$\begin{aligned} E[N(0, n/r) e^{x'r^{1/2}\mathbf{h}}]_{x=0} &= E[N(0, n/\hat{\partial}y) e^{yr + x'r^{1/2}\mathbf{h}}]_{y=0, x=0} \\ &= [N(0, n/\hat{\partial}y) E(e^{yr})]_{y=0} \\ &= [N(0, n/\hat{\partial}y)(1 - 2y)^{-n/2}]_{y=0} \end{aligned} \quad (19)$$

as required.

### The Regression $t$ -Statistic: I

The operational algebra can be illustrated in the case of conventional regression test statistics such as the  $t$ -ratio. We shall proceed under the usual Gaussian assumptions and show how to use the methods to extract the exact distribution theory, the asymptotic distributions and the full asymptotic series expansions. It will be assumed that the reductions from the original data to appropriate sample moments have already been performed.

The sample moments that appear in the usual regression  $t$  statistic are assumed to have been centered and scaled and we write them as  $X \equiv N(0, 1)$ ,  $s \equiv \chi_n^2$ . Then, by independence and in an obvious notation, we have

$$r = \frac{X}{(s/n)^{1/2}} \equiv \frac{N(0, 1)}{(\chi_n^2/n)^{1/2}} \equiv t_n. \quad (20)$$

We write the conditional density of  $r$  given  $s$  and the marginal density of  $s$  as PDF ( $r|s$ )  $\equiv N(0, n/s)$ , PDF ( $s$ ) =  $\chi_n^2(s)$ .

The density of  $r$  is

$$\begin{aligned} \text{PDF}(r) &= \int_{s>0} N(0, n/s) \chi_n^2(s) ds \\ &= \int_{s>0} [N(0, n/\hat{\partial}z) e^{z^2}]_{z=0} \chi_n^2(s) ds \\ &= \left[ N(0, n/\hat{\partial}z) \int_{s>0} e^{z^2} \chi_n^2(s) ds \right]_{z=0} \\ &= [N(0, n/\hat{\partial}z) \text{Lt}(z)]_{z=0} \\ &= [N(0, n/\hat{\partial}z)(1 - 2z)^{-n/2}]_{z=0} \\ &= [N(0, n/2 \hat{\partial}u)(1 - u)^{-n/2}]_{u=0}, \end{aligned} \quad (21)$$

setting  $u = 2z$  so that  $\hat{\partial}z = 2 \hat{\partial}u$ , and

$$\begin{aligned} \text{PDF}(r) &= (2\pi)^{-1/2} [e^{-(r^2/n) \hat{\partial}u} (2/n)^{1/2} \hat{\partial}u^{1/2} (1 - u)^{-n/2}]_{u=0} \\ &= (\pi n)^{-1/2} [e^{-(r^2/n) \hat{\partial}u} \{\Gamma((n+1)/2)/\Gamma(n/2)\} (1 - u)^{-(n+1)/2}]_{u=0} \\ &= \{n^{1/2} B(n/2, 1/2)\}^{-1} (1 + r^2/n)^{-(n+1)/2}, \end{aligned} \quad (22)$$

using (5) above.

The asymptotic distribution as  $n \rightarrow \infty$  can be derived from (21) by noting that  $(1 - u)^{-n/2} \sim e^{-nu/2}$  and then

$$\text{PDF}(r) \sim [N(0, n/2 \hat{\partial}u) e^{-nu/2}]_{u=0} = N(0, 1) \quad (23)$$

directly from (3).

To develop a complete asymptotic series expansion about the limit (23), we start

by writing (21) in the form

$$\text{PDF}(r) = [N(0, 1/\hat{c}v)(1 - 2v/n)^{-n/2}]_{v=0} \quad (24)$$

using the transformation  $u \rightarrow v = nu/2$ . Next, we expand the operand in (24) in the conventional way for large  $n$ , viz.

$$\begin{aligned} (1 - 2v/n)^{-n/2} &= \exp[-(n/2) \ln(1 - 2v/n)] \\ &= \exp\left[+(n/2) \sum_{k=1}^{\infty} (1/k)(2v/n)^k\right] \\ &= e^v \exp\left[\sum_{j=1}^{\infty} 2^j v^{j+1}/n^j(j+1)\right] \\ &= e^v \left[1 + \sum_{k=1}^{\infty} \left(\frac{2}{n}\right)^k \sum_{l=1}^k \frac{v^{k+l}}{l!} \sum^* \{(j_1 + 1) \cdots (j_l + 1)\}^{-1}\right] \end{aligned} \quad (25)$$

where  $\sum^*$  signifies summation over all  $j_1, \dots, j_l (\geq 1)$  for which  $\sum_1^l j_i = k$ . Direct application of the operator in (24) now yields

$$\text{PDF}(r) \sim N(0, 1) + \sum_{k=1}^{\infty} \left(\frac{2}{n}\right)^k \sum_{l=1}^k [N(0, 1/\hat{c}v) e^v v^{k+l}]_{v=0} \frac{1}{l!} \sum^* \{(j_1 + 1) \cdots (j_l + 1)\}^{-1} \quad (26)$$

which by result (15) of the lemma reduces to

$$\text{PDF}(r) \sim N(0, 1) \left[1 + \sum_{k=1}^{\infty} \left(\frac{2}{n}\right)^k \sum_{l=1}^k p_{k+l}(r)\right], \quad (27)$$

where

$$p_{k+l}(r) = \left[\frac{1}{l!} \sum^* \{(j_1 + 1) \cdots (j_l + 1)\}^{-1}\right] \sum_{j=0}^{k+l} \binom{k+l}{j} (-1)^j \left(-\frac{1}{2}\right)_j \left(-\frac{r^2}{2}\right)^{k+l-j} \quad (28)$$

Expression (27) gives a full asymptotic series representation of PDF( $r$ ) about its limit density, the standard normal  $N(0, 1)$ . Note that we use the asymptotic equivalence symbol " $\sim$ " in place of "=", since the series is asymptotic and not convergent. Including terms up to order  $n^{-2}$  in (27), we have

$$\begin{aligned} \text{PDF}(r) \sim N(0, 1) &[1 + (4n)^{-1}(r^4 - 2r^2 - 1) \\ &+ (96n^2)^{-1}(3r^8 - 28r^6 + 30r^4 + 12r^2 + 3) + o(n^{-2})], \end{aligned} \quad (29)$$

which is the same as the expression to this order that was originally found by Fisher (1925).

In a similar way, we can extract an alternate asymptotic series based on Appell



polynomials that was suggested by Dickey (1967). In place of (21) we now use (22) for the calculation of the expansion. Specifically, we have

$$\begin{aligned}
 & [e^{-(r^2/n) \partial u} (1-u)^{-(n+1)/2}]_{u=0} \\
 &= [e^{-(r^2/n) \partial u} \exp[-((n+1)/2) \ln(1-u)]]_{u=0} \\
 &= \left[ e^{-(r^2/n) \partial u} \exp\left[\frac{(n+1)/2}{k=1} \sum_{k=1}^{\infty} u^k/k\right] \right]_{u=0} \\
 &= \left[ e^{-(r^2 h/2) \partial v} \exp\left[\frac{(n+1)/2}{k=1} \sum_{k=1}^{\infty} (v^k/k)(2/(n+1))^k\right] \right]_{v=0}, \\
 & \qquad v = (n+1)u/2, \quad h = (n+1)/n \\
 &= \left[ e^{-(r^2 h/2) \partial v} e^v \left[ 1 + \sum_{k=1}^{\infty} \left(\frac{2}{n+1}\right)^k \sum_{l=1}^k \frac{v^{k+l}}{l!} \sum^* \{(j_1+1) \dots (j_l+1)\}^{-1} \right] \right]_{v=0} \\
 &= e^{-r^2 h/2} \sum_{k=0}^{\infty} \left(\frac{2}{n+1}\right)^k A_k(-hr^2/2), \tag{30}
 \end{aligned}$$

where

$$A_k(x) = \sum_{l=1}^k (1/l!) x^{k+l} \sum^* \{(j_1+1) \dots (j_l+1)\}^{-1} = x^k \sum_{l=1}^k B_{k,l} x^l, \tag{31}$$

say.

In this case the series (30) is convergent for  $r^2 < n$ . The polynomials  $A_k(x)$  are called Appell's polynomials and the coefficients  $B_{k,l}$  may be computed by simple recursion formulas (see Erdélyi (1953) p. 256). Expression (30) leads to a corresponding expansion for the density in terms of Appell polynomials. Dickey (1967) concludes that the accuracy of this expansion is about the same as that of the Edgeworth series (27). The present development helps to show how closely related the two expansions are.

Similar operational methods may be used in the case of the regression  $F$  statistic. Some of the calculations leading to the exact density and distribution function, the  $\chi^2$  asymptotics and full asymptotic series are given in a paper by the author (1987) on fractional operators.

### The Regression $t$ -Statistic: II

The results of the previous section can be derived by conventional methods. The main role of the operational algebra is therefore to simplify the derivations and the final formulas, to bring them all within the same analytical framework and to codify the operations on the probability densities that are induced by the data reductions from sample moments to test statistic.

As a nontrivial application of the techniques, we shall now consider the linear model

$$y = X\beta + u, \quad u \equiv N(0, \Omega) \tag{32}$$

whose error covariance matrix  $\Omega = \Omega(\theta)$  is generally nonscalar and depends on a  $p$ -vector  $\theta \in \Theta \subset \mathbb{R}^p$ , which is variation-independent of the parameter vector  $\beta$  and whose true value is denoted  $\theta_0$ . It will be assumed that  $X$  is nonrandom and of full column rank  $k \leq n$  and  $\Omega(\theta)$  is nonsingular  $\forall \theta \in \Theta$  and analytic. We shall use  $V(\theta) = \Omega(\theta)^{-1}$  to represent the precision matrix. The model is the same as that studied by Rothenberg (1984a, 1984b).

The minimum variance unbiased estimator of  $\beta$  in (32) is given by generalized least squares (GLS) when  $\theta_0$  is known, i.e.

$$\bar{\beta} = [X' V X]^{-1} [X' V y], \quad V = V(\theta_0). \quad (33)$$

When  $\theta_0$  is unknown, it is usual to employ a feasible GLS procedure based on a preliminary estimator  $\hat{\theta}$  of  $\theta_0$ , i.e.  $\hat{\beta} = [X' V(\hat{\theta}) X]^{-1} [X' V(\hat{\theta}) y]$ .

Let  $c'\beta$  be some linear combination of  $\beta$  for some known vector  $c \neq 0$ . Then the corresponding GLS estimate is  $c'\hat{\beta}$  and its asymptotic  $t$ -ratio is

$$t_c = c'(\hat{\beta} - \beta) / \{c' [X' V(\hat{\theta}) X]^{-1} c\}^{1/2}. \quad (34)$$

Our object is to derive a distribution theory for  $c'\hat{\beta}$  and  $t_c$ . To make headway we need to be more explicit about  $\hat{\theta}$ . Since  $\theta$  is usually estimated from the residuals of a first-stage regression on (32), we could let  $\hat{\theta}$  be some function of the residuals from this regression. In fact, it will be more general if we allow  $\hat{\theta}$  to be a function, say

$$\hat{\theta} = \theta(Qu), \quad Qu = (I - X(X' V X)^{-1} X' V)u \quad (35)$$

of the residuals,  $Qu$ , from the GLS regression leading to (33). Here  $Q = Q(\theta_0)$  depends on the true vector  $\theta_0$ . Note that the representation (35) includes all estimators of  $\theta$  that rely on the least squares residuals since  $\hat{u} = (I - P_X)u = (I - P_X)Qu$ .

The formulation (35) also includes the maximum likelihood estimator of  $\theta$ . In this case the log likelihood, after concentrating out  $\beta$ , is

$$\begin{aligned} L(\theta) &= \frac{1}{2} \ln |V(\theta)| - \frac{1}{2} y' Q(\theta)' V(\theta) Q(\theta) y \\ &= \frac{1}{2} \ln |V(\theta)| - \frac{1}{2} y' Q' Q(\theta)' V(\theta) Q(\theta) y. \end{aligned} \quad (36)$$

Since  $L(\theta)$  depends on  $y$  only through  $Qy$ , so too does its optimum and hence (35) applies in this case also.

We now write the estimator  $\hat{\beta}$  in terms of  $\bar{\beta}$  and  $Qu$  as  $\hat{\beta} = \bar{\beta} + (\hat{\beta} - \bar{\beta}) = \bar{\beta} + d$ , where  $d = [X' V(\hat{\theta}) X]^{-1} X' V(\hat{\theta}) Qu = d(Qu)$ , say, is statistically independent of  $\bar{\beta}$ . Let  $r = c'\hat{\beta} = c'\bar{\beta} + c'd$  and we have the conditional density

$$\text{PDF}(r | Qu) = N(c'\bar{\beta} + c'd, c'(X' V X)^{-1} c). \quad (37)$$

Next  $Qu$  is singular  $N(0, QV^{-1}Q)$  and using  $P(\cdot)$  to represent the probability measure

of this distribution, we deduce that

$$\begin{aligned}
 \text{PDF}(r) &= \int N(\mathbf{c}'\boldsymbol{\beta} + \mathbf{c}'\mathbf{d}, \mathbf{c}'(X'VX)^{-1}\mathbf{c}) dP(Qu) \\
 &= \int [N(\mathbf{c}'\boldsymbol{\beta} + \mathbf{c}'\mathbf{d}(\partial w), \mathbf{c}'(X'VX)^{-1}\mathbf{c}) e^{w'Qu}]_{w=0} dP(Qu) \\
 &= \left[ N(\mathbf{c}'\boldsymbol{\beta} + \mathbf{c}'\mathbf{d}(\partial w), \mathbf{c}'(X'VX)^{-1}\mathbf{c}) \int e^{w'Qu} dP(Qu) \right]_{w=0} \\
 &= [N(\mathbf{c}'\boldsymbol{\beta} + \mathbf{c}'\mathbf{d}(\partial w), \mathbf{c}'(X'VX)^{-1}\mathbf{c}) e^{w'QV^{-1}Qw/2}]_{w=0}. \tag{38}
 \end{aligned}$$

Note that in this formula  $\mathbf{d}(\partial w)$  and hence  $N(\mathbf{c}'\boldsymbol{\beta} + \mathbf{c}'\mathbf{d}(\partial w), \mathbf{c}'(X'VX)^{-1}\mathbf{c})$  are analytic functions of the differential operator  $\partial w$  and can be interpreted according to the definition (6).

Expression (38) shows the distribution of  $\hat{\boldsymbol{\beta}}$  to be a form of normal mixture. There is no linear term in the exponent of the operand  $\exp(w'QV^{-1}Qw/2)$  since this is the Laplace transform of the distribution of  $Qu$ , whose mean is zero. Upon replacement of this operand with the approximation  $\exp(0'w) = 1$  we see that (38) reduces to

$$N(\mathbf{c}'\boldsymbol{\beta}, \mathbf{c}'(X'VX)^{-1}\mathbf{c}), \tag{39}$$

which is the usual first-order asymptotic approximation to the distribution of  $\mathbf{c}'\hat{\boldsymbol{\beta}}$ . Higher order approximations may also be obtained. Note that when  $\hat{\theta} = \theta(Qu)$  is taken to be an even function of  $Qu$ , as it is in most practical situations, it is easy to show from (38) that the next term in the expansion leads only to an adjustment in the asymptotic variance term in (39). The normal approximation itself is retained. This corresponds with the main conclusion of Rothenberg (1984a).

The distribution of the  $t$ -ratio  $t_c$  in (34) may be handled in the same way as  $\hat{\boldsymbol{\beta}}$ . With  $t_c$  we have the additional dependence of the denominator on  $\hat{\theta}$ . However, this presents no further complication in the operational algebra. Following the same steps as those employed in deriving (38), we find

$$\text{PDF}(t) = [N(\mathbf{c}'\mathbf{d}(\partial w), \mathbf{c}'(X'VX)^{-1}\mathbf{c}/\mathbf{c}'(X'V(\partial w)X)^{-1}\mathbf{c}) e^{w'QV^{-1}Qw/2}]_{w=0}. \tag{40}$$

This formula can be interpreted as an extension to the general regression case of formula (21) given earlier for the distribution of the usual  $t$ -ratio. In that case the error covariance matrix  $\Omega = \sigma^2 I$  was scalar. Writing  $V(\hat{\theta}) = (1/\hat{\sigma}^2)I$  with  $\hat{\sigma}^2 = \hat{u}'\hat{u}/n$  and  $n = T - k$ , we see that  $\mathbf{d}(\partial w) = 0$ ,  $V(\partial w) = (1/\partial w' \partial w)I$  and formula (40) above reduces to

$$[N(0, n\sigma^2/\partial w' \partial w) e^{\sigma^2 w' Qx w/2}]_{w=0}, \tag{41}$$

where  $Q_x = I - P_x$ . Let  $C$  be an orthogonal matrix for which  $C'Q_x C = \text{diag}(I_n, 0)$ . Transform  $w \rightarrow z = \sigma C'w$  and note that the operators transform according to the relation  $\partial w = \sigma C \partial z$ .

Under this transformation (41) becomes

$$[N(0, n/\partial z' \partial z) e^{z'iz_1/2}]_{z=0} = [N(0, n/\partial z'_1 \partial z_1) e^{z'iz_1/2}]_{z_1=0}, \quad (42)$$

where  $z$  is partitioned as  $(z'_1, z'_2)'$ , conformably with the block diagonal decomposition of  $C'Q_X C$ . The expression on the right-hand side of (42) now simplifies by a change of dummy variables to  $[N(0, n/\partial u)(1 - 2u)^{-n/2}]_{u=0}$  as in result (16) of the lemma in the second section. The density given in (42) therefore reduces to that of the usual  $t$ -ratio in this case.

In general, of course, the symbolic representations (38) and (40) rely on the form of the dependence of  $V(\theta)$  on  $\theta$ . In models with heterogeneity, the dependence is straightforward. In finite-order autoregressive error models,  $V(\theta)$  has the usual finite-band matrix format associated with the inverse of the covariance matrix. For more general ARMA models the dependence is more complex but formulas are known (Zinde-Walsh 1988).

### Conclusion

The methods introduced in this paper belong to a class of operational techniques that have been used for some time by mathematicians in studying solutions to systems of partial differential equations. The literature in this field is vast; but it is also rather abstract, as is evident in the works of Hormander (1971), Treves (1982) and Dieudonné (1988). Little attention seems yet to have been given to the development of practical rules for working with nonlinear functions of operators in specific cases of interest. Nevertheless, it would appear from the regression examples studied here that the methods offer some prospect for dealing with rather complex distributional problems in an economical and efficient way. Indeed, as is apparent from the previous section, the regression estimator and  $t$ -statistic in the general linear model with a nonscalar covariance matrix are as easy to handle with the operational algebra as are the OLS estimator and  $t$ -ratio in the classical regression model.

### Notes

My thanks go to a referee for helpful comments, to Glenna Ames for her talents in word-processing this essay and to the NSF for research support.

### References

- Basman, R. L. (1961). "Note on the exact finite sample frequency functions of generalized classical linear estimators in two leading overidentified cases," *Journal of the American Statistical Association*, 55, 619-36.
- Bergstrom, A. R. (1962). "The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume," *Econometrica*, 30, 480-90.
- Dieudonné, J. (1988). *Treatise on Analysis*, volume 8. New York: Academic Press.
- Dickey, J. M. (1967). "Expansions of  $t$  densities and related complete integrals," *Annals of Mathematical Statistics*, 503-10.
- Erdélyi, A. (1953). *Higher Transcendental Functions*, volume 3. New York: McGraw-Hill.
- Fisher, R. A. (1925). "Expansion of Student's integral in powers in  $n^{-1}$ ," *Metron*, 5, 109-12.

- GAUSS (1989). *Gauss Version 2.0*. Kent: Aptech Systems Inc.
- Gelfand, I. M. and G. E. Shilov (1964). *Generalized Functions*, volume 1. New York: Academic Press.
- Hormander, L. (1971). "Fourier integral operators," *Acta Mathematica*, 127, 79–183.
- Miller, K. S. (1974). "The Weyl fractional calculus," in B. Ross (ed.) *Fractional Calculus and its Applications*. Berlin: Springer-Verlag.
- Phillips, P. C. B. (1984). "The exact distribution of the Stein-rule estimator," *Journal of Econometrics*, 25, 123–31.
- Phillips, P. C. B. (1985). "The exact distribution of the SUR estimator," *Econometrica*, 53, 745–56.
- Phillips, P. C. B. (1986). "The exact distribution of the Wald statistics," *Econometrica*, 54.
- Phillips, P. C. B. (1987). "Fractional matrix calculus and the distribution of multivariate tests," in I. B. MacNeil and G. J. Umphrey (eds) *Time Series and Econometric Modeling*. Dordrecht: Reidel Publishing Company.
- Richards, D. (1984). "Hyperspherical models, fractional derivatives and exponential distributions on matrix spaces," *Sankhya A*, 45, 155–65.
- Rothenberg, T. J. (1984a). "Approximate normality of generalized least squares estimates," *Econometrica*, 52, 811–26.
- Rothenberg, T. J. (1984b). "Hypothesis testing in linear models when the error covariance matrix is nonscalar," *Econometrica*, 52, 827–42.
- Treves, F. (1982). *Pseudodifferential and Fourier Integral Operators*, volumes 1 & 2. New York: Plenum Press.
- Zinde-Walsh, V. (1988). "Some exact formulae for autoregressive moving average processes," *Econometric Theory*, 4, 384–402.