Econometrica, Vol. 50, No. 3 (May, 1982)

NOTES AND COMMENTS

A COMPUTATIONALLY EFFICIENT QUADRATURE PROCEDURE FOR THE ONE-FACTOR MULTINOMIAL PROBIT MODEL

BY J. S. BUTLER AND ROBERT MOFFITT[1]

A PROBLEM OF ESTIMATION that has long confronted many economists is the difficulty of estimating the parameters of equations with limited dependent variables on cross-section time-series (i.e., panel) data. While there are widely available packaged computer programs for estimating either (a) cross-section probit and Tobit models or (b) simple permanent-transitory, random-effects panel models with continuous dependent variables, there are no available computationally feasible methods of combining these two models. This is because the likelihood function that arises in such a combined model contains multivariate normal integrals whose evaluation is quite difficult, if not impossible, with conventional approximation methods. There is a widespread feeling among those working in the area that one possible method of evaluation, the use of quadrature techniques, is in principle possible but is in practice computationally too burdensome to consider (e.g., Albright et al. [**2**, p. 13]; Hausman and Wise [**6**, p. 12]). In this note we point out that this is true only of standard quadrature techniques such as trapezoidal integration or its improved variants; Gaussian quadrature, on the other hand, is extremely efficient and is well within the bounds of computational feasibility on modern computers. In what follows, we state the nature of the integrals that need to be evaluated, provide a brief exposition of Gaussian quadrature, and provide a numerical illustration of its use in estimating a one-factor multinomial probit model.

Assume we have the following panel probit model:

$$Y_{it}^* = X_{it}\beta + \epsilon_{it} \qquad (i = 1, \ldots, N; t = 1, \ldots, T),$$

$$Y_{it} = \begin{cases} 1 & \text{if } Y_{it}^* \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where i indexes individuals, t indexes time periods, X is a vector of independent variables, and β is a vector of corresponding coefficients. Assume that the disturbances are generated by the permanent-transitory process $\epsilon_{it} = \mu_i + \nu_{it}$, where $\epsilon_{it} \sim N(0, \sigma^2)$ and ρ is the correlation between successive disturbances for the same individual. The log-likelihood function for the problem is $L = \sum_{i=1}^{N} \log[\text{prob}(Y_{i1}, \ldots, Y_{iT})]$, where

$$\text{(1)} \qquad \text{prob}(Y_{i1}, \ldots, Y_{iT}) = \int_{a_{i1}}^{b_{i1}} \cdots \int f(\epsilon_{i1}, \ldots, \epsilon_{iT})\, d\epsilon_{iT} \ldots d\epsilon_{i1}$$

and $a_{it} = -X_{it}\beta$ and $b_{it} = \infty$ if $Y_{it} = 1$, and $a_{it} = -\infty$ and $b_{it} = -X_{it}\beta$ if $Y_{it} = 0$, and $f(.)$ is the normal density function. The standard difficulty in this problem is the evaluation of the T-fold integrals in equation (1). Since the random components are independent conditional upon the permanent component, the integral can be simplified by condition-

[1]The authors would like to thank Mathematica Policy Research for subsidizing the research reported herein. Comments from Randall Brown and Timothy Carr as well as from two anonymous referees are much appreciated.

ing on the permanent component:

(2) $$\operatorname{prob}(Y_{i1}, \ldots, Y_{iT})\operatorname{prob}(Y_{i1}, \ldots, Y_{iT})$$
$$= \int_{a_{i1}}^{b_{i1}} \cdots \int_{a_{iT}}^{b_{iT}} \int_{-\infty}^{\infty} f(v_{i1} | \mu_i) f(\mu_i)\, d\mu_i\, dv_{iT} \cdots dv_{i1}$$
$$= \int_{-\infty}^{\infty} f(\mu_i) \prod_{t=1}^{T} [F(b_{it} | \mu_i) - F(a_{it} | \mu_i)]\, d\mu_i$$

where $F(\ )$ is the normal cumulative distribution function (cdf).[2] Thus, the expression can be reduced to a single integral whose integrand is a product of one normal density and T differences of normal cdf's for which highly accurate approximations are available. Nevertheless, even the evaluation of the single integral in equation (2) is extremely burdensome using conventional quadrature procedures such as trapezoidal integration or its variants such as Romberg integration.[3] Gaussian quadrature, on the other hand, is a much more sophisticated procedure that requires the evaluation of the integrand at many fewer points in the domain of μ, thus achieving gains in computational efficiency of several orders of magnitude. The formula for the evaluation of the necessary integral is the Hermite integration formula $\int_{-\infty}^{\infty} e^{-Z^2} g(Z)\, dZ = \sum_{j=1}^{G} w_j g(Z_j)$, where G is the number of evaluation points, w_j is the weight given to the jth evaluation point, and $g(Z_j)$ is $g(Z)$ evaluated at the jth point of Z.[4] This formula is appropriate to our problem because the normal density f in equation (2) contains a term of the form $\exp(-Z^2)$, and the function $g(Z)$ is, in our case, the product of T univariate cdf's.

The key question for computational feasibility is the number of points at which the integrand must be evaluated for accurate approximation. Several evaluations of the integral using four periods of arbitrary values of the data and coefficients on six right-hand-side variables showed us that even two-point integration is highly accurate.[5] Table I provides estimates of a fertility equation (dependent variable = 1 if a birth in year t, 0 if not) on a sample of 1550 women with a maximum of 11 periods each, drawn from the Young Women's cohort of the National Longitudinal Survey. The algorithm of Berndt et al. [**3**] was used for maximization of the likelihood function.[6] As the table shows, some

[2] This factorization is "periodically and independently rediscovered" according to Gupta [**5**, p. 800] and has recently been discussed by Heckman [**7**]. See Heckman's paper for an explicit representation of the conditional cdf's equation (2).

[3] For example, Heckman and Willis [**8**] used the extremely expensive trapezoidal technique. These methods approximate integrals such as that in (2) by a polynomial in the integrand evaluated at several equally spaced intervals in the domain of the integrating variable μ, which is quite expensive because the integrand must generally be evaluated at many points for the approximation to be of acceptable accuracy. See Pennington [**10**, 247–251] for a discussion.

[4] See Stroud and Secrest [**11**, p. 22] and Pennington [**10**, p. 260]. Hermite integration can evaluate the integral $\int_{-\infty}^{\infty} e^{-Z^2} g(Z) dZ$ exactly with P points if $g(Z)$ is a polynomial of degree less than $2P-1$. Romberg integration requires $2^P - 1$ points to attain the same degree of accuracy.

[5] Using synthetic data and arbitrary coefficients, and with $X\beta$ ranging from .60 to .75 in four periods of hypothetical data, the value of the integral for 2, 3, 4, and 5 evaluation points was .31735585, .31734161, .31734174, and .31734174, respectively. The points and weights for the calculations are available from several easily accessible sources: Stroud and Secrest [**11**], Abramowitz and Stegun [**1**, p. 924]. We also included an Appendix table in an earlier version of this paper giving the points and weights for two-point to five-point integrations. This is available upon request.

[6] The starting values used were obtained by estimating the equation on only five periods of data, but using the approximation of Clark [**4**] for the evaluation of the integrals. Probit starting values could also have been used. The Clark approximation was tested on the eleven-period data and was found to be extremely inaccurate, even for the one-factor error structure assumed here. For example, the area under the 11-fold surface summed to 1.9 instead of 1.0. The inaccuracy was reduced the fewer the number of periods; for 5 the Clark approximation appeared to be reasonably accurate in the present problem.

TABLE I

COEFFICIENT ESTIMATES OF A MULTINOMIAL PROBIT FERTILITY EQUATION USING GAUSSIAN QUADRATURE MAXIMUM LIKELIHOOD

Variables[a]	Number of Evaluation Points 2	3	4	5
Constant	0.198006 (0.531493)	0.579574 (0.636880)	0.681029 (0.739860)	0.895984 (0.739124)
Race	0.435691* (0.048866)	0.391777* (0.062498)	0.372754* (0.066584)	0.399289* (0.067595)
Education	− 0.087276* (0.011250)	− 0.104133* (0.013863)	− 0.101742* (0.016208)	− 0.107280* (0.016285)
Coh	− 0.023634* (0.009520)	− 0.028942* (0.011232)	− 0.030927* (0.012773)	− 0.034811* (0.012805)
Wealth	− 0.031125 (0.049793)	0.058309 (0.061394)	0.020812 (0.064974)	0.041226 (0.066930)
Oths	0.028244 (0.024087)	0.043871** (0.026354)	0.048919** (0.027525)	0.049346** (0.027569)
Time	0.607854* (0.028321)	0.612613* (0.029532)	0.609026* (0.029249)	0.614247* (0.029550)
Time Squared	− 0.040195* (0.002711)	− 0.040179* (0.002856)	− 0.039651* (0.002817)	− 0.040261* (0.002850)
Rho	0.298808* (0.011360)	0.322650* (0.015322)	0.326820* (0.015920)	0.339101* (0.016491)
CPU time (seconds)[b]	81.8	114.4	116.8	169.6

NOTE: Standard errors in parentheses.
*Significant at the 5 per cent level of confidence.
**Significant at the 10 per cent level of confidence.
[a] Race = 1 if nonwhite, 0 if not; Education = years of schooling; Coh = year of birth (e.g., "48" for 1948); Wealth = discounted present value of lifetime family income other than wife's earnings (1967 dollars); Oth = number of adults in family other than husband and wife; Time = duration of marriage in years. Taken from Moffitt [9].
[b] Hermite points are symmetric about zero and include zero if there is an odd number of them. For purposes of calculation, then, $2k-1$ and $2k$ evaluation points are virtually equivalent, having k different absolute values. CPU times mainly rise in going from 2 to 3, 4 to 5, etc., evaluation points.

coefficients change substantially, the more accurate the approximation of the likelihood function. However, those that change much at all are insignificant even at the lowest, two-point integration; no hypothesis test at the 5 per cent level of confidence would come out differently at the four-point evaluation and at the two-point evaluation. Going from 2 to 3 point integration changes the "significant" coefficients (those on "Oths" and the six with t values above 2) an average of 16.2 per cent; going from 3 to 4 changes them an average of 4.1 per cent; and going from 4 to 5 changes them an average of 3.6 per cent.[7]

[7] The average change in going from 3 to 5 points is 6.1 per cent. The likelihood function and its derivatives are being estimated more accurately with more points, but there is no simple link between this accuracy and the direction of change in the estimated coefficients. Four, six, and eight Hermite points replace 15, 31, and 127 trapezoids.

Thus, those coefficients significant at the two-point evaluation are changed very little by the increase in accuracy. Therefore, the evidence shows clearly that a two-point evaluation is quite satisfactory for hypothesis testing, and probably also for significant coefficient values. The CPU times (for an IBM 370) shown at the bottom are well within financial feasibility at most academic and nonacademic institutions.

We conclude with several points. First, in the context of a maximization algorithm, accuracy could be increased and costs reduced by raising the number of evaluation points as the likelihood function approaches its optimum. Second, a more general point is that the technique is clearly applicable to other limited-dependent-variable models such as single-bound Tobit, double-bound Tobit, and others that are currently proliferating. The modification required in each case is to replace the cdf's in equation (2) with whatever the appropriate cross-section analogue is (e.g., simple probability density functions for Tobit observations above the limit). Third, though we have not tested it ourselves, two-fold-integration by Gaussian methods may also be within the bounds of feasibility; this would allow the estimation of two-factor models as well. Fourth, the technique is applicable to the cross-section multiple-indicator model as well. That the dependent variables are limited in these models is usually ignored.[8]

Mathematica Policy Research, Inc., Princeton, New Jersey

and

Rutgers University and Mathematica Policy Research, Inc., Princeton, New Jersey

Manuscript received November, 1980; revision received June, 1981.

[8] A copy of the deck used to estimate the equations in Table I is available from the authors at reproduction cost for two years from the date of publication.

REFERENCES

[1] Abramowitz, M., and I. Stegun: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards Applied Mathematics Series No. 55. Washington, D.C.: U.S. Government Printing Office, 1964.

[2] Albright, R., S. Lerman, and C. Manski: "Report on the Development of an Estimation Program for Mutinomial Probit Model," Report prepared for the Federal Highway Administration, 1977.

[3] Berndt, E., B. Hall, R. Hall, and J. Hausman: "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3(1974), 653–665.

[4] Clark, C: "The Greatest of a Finite Set of Random Variables," *Operations Research*, 9(1964), 145–162.

[5] Gupta, S.: "Probability Integrals of Multivariate Normal and Multivariate *t*," *Annals of Mathematical Statistics*, 34(1963), 792–838.

[6] Hausman, J., and D. Wise: "AFDC Participation: Measured Variables or Unobserved Characteristics, Permanent or Transitory," mimeographed, Department of Economics, M.I.T., 1979.

[7] Heckman, J.: "Statistical Models for Discrete Panel Data," in *The Econometrics of Panel Data*, ed. by D. McFadden and C. Manski. Cambridge: MIT Press, 1981.

[8] Heckman, J., and R. Willis: "Estimation of a Stochastic Model of Reproduction," in *Household Production and Consumption*, ed. by N. Terleckyj. New York: Columbia University Press, 1976.

[9] Moffitt, R.: "Life Cycle Profiles of Fertility: A State-Dependent Multinomial Probit Model," mimeo, Rutgers University, 1980.

[10] Pennington, R.: *Introductory Computer Methods and Numerical Analyses* (second edition). London: The McMillian Company, 1970.

[11] Stroud, A., and D. Secrest: *Gaussian Quadrature Formulas*. Englewood Cliffs: Prentice-Hall, 1966.