Journal of Econometrics 28 (1985) 127–134. North-Holland

COMPUTATION IN DURATION MODELS WITH HETEROGENEITY

Donald M. WALDMAN

University of Colorado, Boulder, CO 80309, USA

This paper examines the use of Hermite polynomial quadrature to evaluate certain integrals in the maximum likelihood estimation of duration models.

1. Introduction

Computation of the likelihood in duration data models with unobservable individual effects poses special problems. Kiefer (1983) has provided an unusual and effective method of performing these calculations, for certain stochastic assumptions. The 'random effects' specification implies that the likelihood is composed of terms which are the expectation of a function of the effect, $\mathrm{E}[f(z)]$. If z is assumed to be standard normal, the usual case, and the model is a duration data model with certain hazards (for example, those admitting to a closed form, such as the Weibull), then the moments of $\log f(z)$ are calculable. Kiefer suggests approximating $\mathrm{E}[f(z)]$ by the expectation of the truncated expansion of $\exp[\log f(z)]$,

$$\begin{aligned}\mathrm{E}[f(z)] &= \mathrm{E}[\exp(\log f(z))] \\ &\doteq \mathrm{E}\left[1+\sum_{i=1}^{m}(\log f(z))^{i}/i!\right] \\ &= 1+\sum_{i=1}^{m}\mu'_i/i!,\end{aligned}$$

where μ'_i is the ith raw moment of $\log f(z)$ and m is the number of terms in the expansion. This approximation can be made as accurate as required for successful maximum likelihood (ML) estimation by increasing m.

For other models with heterogeneity the moments of the resulting $\log f(z)$ cannot be expressed algebraically in convenient form. Another method of calculation is therefore necessary. In this paper we propose the use of a variation of Gaussian quadrature. This type of numerical integration has been recently suggested by Butler and Moffitt (1982) for computation in multinomial probit models. The method is examined for accuracy and speed in the ML estimation of the parameters of a model of technological diffusion. Then, with repeated evaluation of individual integrals, we demonstrate its superiority in those models where the expansion above *is* applicable.

2. The model

Here we consider the adoption by individual firms of a menu of technological innovations available to them. Let t_{ij}^* represent the desired length of time to wait before the ith firm adopts the jth innovation. Left-censoring occurs because a subset of firms all observed to have adopted the innovation immediately upon its introduction were unlikely to be equally desirous of doing so.[1] Right-censoring occurs because at the time of the survey some firms may not (yet) have adopted. Define observed adoption time as

$$\begin{aligned} t_{ij} &= 0 \quad \text{if} \quad t_{ij}^* \le 0, \\ &= t_{ij}^* \quad \text{if} \quad 0 < t_{ij}^* \le L_j, \\ &= L_j \quad \text{if} \quad L_j < t_{ij}^*, \end{aligned} \tag{1}$$

where L_j is the length of time the jth innovation has been available at the time of the survey. Desired adoption time is assumed to be a function of characteristics of the firms and innovations, distinguished by their subscripts, plus additive random disturbance,

$$t_{ij}^* = g(x_i, z_j; \theta) + \varepsilon_{ij}, \qquad i = 1, \dots, T, \quad j = 1, \dots, N, \tag{2}$$

where θ is a vector of unknown parameters. If the strength of desire to adopt can be adequately captured in the function g, the ε_{ij} can be interpreted as measurement error and assumed to follow the normal distribution,[2]

$$\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2), \quad \forall i, j. \tag{3}$$

Since there are two dimensions to our data, many of the issues involved in estimating with time-series cross-section data arise here. For industry data where there are typically many firms and few innovations, it would seem prudent to allow for innovation-specific effects (estimate conditionally on the

[1] For a complete discussion of the model, including the somewhat unusual feature of left-censored observations, see Waldman and Lee (1983). For other examples of left-censoring, see Leiderman et al. (1973) and Turnbull (1974).

[2] For other examples of the use of the normal distribution, see Schmee and Hahn (1979).

innovation effect) while including a random component to allow firms to differ with respect to their average (across innovations) adoption duration. Disaggregating the disturbance in (3) gives

$$\varepsilon_{ij} = \mu_i + \nu_{ij}, \tag{4}$$

where we assume $\mu_i \sim \mathrm{N}(0, \sigma_\mu^2)$, ν_{ij} is white noise, and $\mathrm{E}(\mu_i \nu_{ij}) = 0, \forall i, j$.

3. The likelihood and its computation

The likelihood function for the model of eqs. (1)–(4) is the product of T terms, each an N-variate multinormal probability. The multinormal density must be considered because $\mathrm{E}(\varepsilon_{ij}\varepsilon_{ik}) = \sigma_\mu^2$ for $j \neq k$. We may condition on the cause of this correlation, so that the likelihood becomes the product of T terms, each the integral of one normal density and N normal densities or cdf's, depending upon the pattern of adoption for that firm. For example, suppose $N = 3$ (three innovations to consider) and a firm adopts one innovation as soon as it is offered ($t_{i1} = 0$), adopts another after some delay ($t_{i2} = t_{i2}^*$), and, as of the date of the survey, had not adopted the third innovation ($t_{i3} = L_3$). Then the factor in the likelihood corresponding to this firm would be

$$\Pr\{\varepsilon_{i1} < -g_{i1}, \varepsilon_{i2} = t_{i2} - g_{i2}, \varepsilon_{i3} > L_3 - g_{i3}\}$$

$$= \frac{1}{\sigma_\mu} \int_{-\infty}^{\infty} \phi(\mu_i/\sigma_\mu)$$

$$\times \Pr\{\nu_{i1} < -\mu_i - g_{i1}, \nu_{i2} = -\mu_i + t_{i2} - g_{i2}, \nu_{i3} > -\mu_i + L_3 - g_{i3}\}\, \mathrm{d}\mu_i$$

$$= \frac{1}{\sigma_\mu} \int_{-\infty}^{\infty} \phi(\mu_i/\sigma_\mu)\, \Phi[(-\mu_i - g_{1i})/\sigma]\, \frac{1}{\sigma}\phi[(-\mu_i + t_{i2} - g_{i2})/\sigma]$$

$$\times \Phi[-(-\mu_i + L_3 - g_{i3})/\sigma]\, \mathrm{d}\mu_i, \tag{5}$$

where $g_{ij} = g(x_i, z_j; \theta)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cdf, respectively. Using conventional numerical integration, computation of (5) is slow, requiring many points at which to evaluate the integrand to insure sufficient accuracy.[3] Gaussian quadrature is a powerful tool that can reduce the number of integration points impressively. From Abramowitz and Stegun

[3] In some other applications, twenty-five points were sufficient. In our experience, different models have required as many as forty to fifty points.

(1972, p. 890) we make use of

$$\int_{-\infty}^{\infty} e^{-x^2} I(x)\,dx = \sum_{i=1}^{n} w_i I(x_i) + R_n, \tag{6}$$

where $I(x)$ is that part of the integrand in (5) with e^{-x^2} factored out, x_i is the ith zero of the Hermite polynomial $H_n(x)$, n is the number of grid points, and the weights w_i are

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_i)]^2}. \tag{7}$$

(R_n is the remainder term of the expansion.) Tables of x_i and w_i are provided for $n = 2$–10 (1), 12, 16, and 20 in the above cited reference.

To demonstrate this method of quadrature and to investigate the effect of various values of n, we maximized the likelihood for a model where a cross-section of 300 hospitals consider adopting five medical innovations.[4] Thus $T = 300$ and $N = 5$. Individual constants are estimated for each innovation. Explanatory variables include hospital size (*ADJPD5*), an indicator variable which takes a value of one if the observation is a for-profit hospital (*PROFIT*), the severity of the climate of the hospital which is interacted with only the indicator variable corresponding to the energy management innovation (*TMPDIF*), and the number of births at the hospital which is interacted with only the indicator variable for a fetal-monitoring innovation (*BRTHS5*). Table 1 presents parameter estimates, asymptotic t-ratios, and other information for maximum likelihood estimation with the number of grid points varying from two to ten.[5]

The results clearly demonstrate the power of the computational method. From a minimum of two, parameter estimates are only slightly affected by a five-fold increase in the number of grid points. The only appreciable change comes when moving from two to four points, with some changes of approximately four percent. For additional grid points, changes are approximately of the order 10^{-1} percent. Estimated asymptotic t-ratios are well-behaved, except for those for *TMPDIF* and *INNOV5*, which change somewhat erratically.[6] The maximized value of the log-likelihood has essentially stopped changing for $n \geq 8$.

[4] The data are from a 1980 survey conducted by the American Hospital Association. See Romeo et al. (1982).

[5] Starting values were the same for each estimation: zero's for the slope coefficients, the sample means for the innovation-specific constants, which were 6.157, 5.593, 16.283, 5.520, and 7.157, and one's for the two variance parameters.

[6] Differences in t-ratios would *not* alter an evaluation of statistical significance. These estimates appear to have stabilized for $n \geq 6$.

Table 1

Maximum likelihood estimates of truncated normal duration of the innovation adoption model.

Variable name	Number of grid points 2	4	6	8	10
ADJPD5	−0.90238	−0.89298	−0.89316	−0.89240	−0.89260
	(9.74)	(9.58)	(9.93)	(9.21)	(9.25)
PROFIT	−0.49949	−0.51385	−0.51739	−0.51257	−0.51520
	(1.13)	(1.08)	(1.05)	(1.05)	(1.04)
TMPDIF	−0.06248	−0.06053	−0.06144	−0.06305	−0.06273
	(7.67)	(3.49)	(2.97)	(1.97)	(2.05)
BRTHS5	−0.00337	−0.00338	−0.00338	−0.00333	−0.00333
	(2.06)	(2.08)	(2.09)	(2.04)	(2.03)
INNOV1	7.27879	7.28409	7.28309	7.28030	7.28127
	(37.60)	(38.77)	(39.08)	(38.79)	(38.72)
INNOV2	6.39854	6.40067	6.40122	6.40238	6.40236
	(42.12)	(41.74)	(42.46)	(41.53)	(41.22)
INNOV3	17.63405	17.63699	17.63676	17.63715	17.63748
	(127.50)	(125.74)	(128.10)	(123.09)	(121.84)
INNOV4	6.32851	6.33052	6.33147	6.33098	6.33117
	(42.69)	(41.43)	(42.36)	(41.05)	(40.96)
INNOV5	10.81301	10.72781	10.76681	10.83831	10.82363
	(28.58)	(13.69)	(11.19)	(7.49)	(7.84)
σ_v^2	0.56905	0.58840	0.58715	0.58292	0.58318
	(8.26)	(7.96)	(6.80)	(6.44)	(6.29)
σ_μ^2	2.36998	2.36607	2.36618	2.36698	2.36698
	(59.77)	(56.93)	(54.80)	(55.84)	(54.95)
Log-L	−3276.7019	−3276.3853	−3276.3893	−3276.3922	−3276.3921
CPU time (in seconds)	50.55	103.98	157.80	200.49	247.99

Reported CPU times are very generous. In each case the DFP algorithm had converged (with respect to changes in parameter estimates and likelihood) after fifteen to seventeen iterations, but the algorithm was allowed to continue until twenty-five iterations had been completed.[7] Therefore actual times to convergence might be closer to 60 to 70% of those reported in the table.

We have demonstrated the speed and accuracy of Gaussian quadrature for the evaluation of the duration data likelihood when there is heterogeneity. This method is general with respect to stochastic assumption – our example involved normality, which does not admit to a closed form. We now turn to an

[7]First derivatives of the likelihood at the converged values of the parameter estimates were usually in the range $10^{-4} - 10^{-1}$. The GQOPT subroutine package was employed.

examination of the efficiency of this method relative to the expansions presented in Kiefer (1983) for certain stochastic assumptions where a closed form exists.

4. A comparison of methods

If the function $I(x)$ in (6) takes the form $I(x)=\exp(ax+be^{cx})$, then the moments about zero of $\log I(x)$ are

$$\mu'_r=\sum_{i=0}^{r}\left\{\binom{r}{i}a^{r-i}b^i e^{i^2c^2/2}\left[\sum_{j=0}^{[(r-i)/2]}\binom{r-i}{2j}(ic)^{r-i-2j}\left(\frac{(2j)!}{2^j 2!}\right)\right]\right\}, \quad (8)$$

where $[(r-i)/2]$ is the greatest integer less than or equal to $(r-i)/2$ and 0^0 is set to one in the second summation. The integral is then approximated by the expansion

$$(2\pi)^{-1/2}\int_{-\infty}^{\infty} I(x)e^{-1/2x^2}dx \doteq 1+\sum_{i=1}^{m}\mu'_i/i!, \quad (9)$$

or by the expansion with central moments

$$(2\pi)^{-1/2}\int_{-\infty}^{\infty} I(x)e^{-1/2x^2}dx \doteq e^{\mu'_1}\left[\sum_{i=1}^{m}\mu_i/i!+1\right], \quad (10)$$

where the moments can be calculated from (8) by

$$\mu_i=\sum_{j=0}^{i}(-1)^j\binom{i}{j}\mu_1'^{j}\mu'_{i-j}.$$

[See Kiefer (1983, eqs. 6–8).]

In table 2 we report an example calculation of an integral of this form by the two expansions (9), (10) after calculation of the moments by (8), and by quadrature. The method of quadrature converges to seven places of accuracy after the decimal point when the number of grid points (n) reaches seven. The expansion methods each require fifteen terms (m). But for feasibility for use as part of a function maximization algorithm, time and not the number of points is the relevant consideration. Evaluation of the integral by quadrature requires six elementary operations [two exponentiations, three multiplications, and one addition to compute $I(x)$] in the function, times two [one multiplication, one addition] for each grid point, for a total of $12n$ calculations. The calculation of the rth moment in (8) requires approximately

$$10(r+1)+\tfrac{7}{4}(r+1)(r+3)$$

Table 2

Example calculation of the integral.[a]

m, n[b]	Expansion (9)	Expansion (10)	Quadrature
2	1.1459239	1.0305905	1.0538379
3	1.0004764	1.0222021	1.0874061
4	1.1082490	1.0832517	1.0882979
5	1.0746221	1.0798160	1.0882038
6	1.0917508	1.0880818	1.0881973
7	1.0865691	1.0873843	1.0881976
8	1.0886380	1.0882271	—
9	1.0880383	1.0881333	—
10	1.0882402	1.0882027	—
11	1.0881845	1.0881933	—
12	1.0882010	1.0881981	—
13	1.0881967	1.0881974	—
14	1.0881979	1.0881977	—
15	1.0881976	1.0881976	—

[a] Each entry is an approximation of $(2\pi)^{-1/2}\int_{-\infty}^{\infty}\exp(1-0.3\,e^{0.3x})\,e^{-1/2x^2}\,dx$.
[b] Number of terms in expansion (m in text), or number of grid points (n in text).

calculations [from an examination of (8)], exclusive of combinations and factorials (which were put in a table, as suggested by Professor Kiefer). For the expansion in (9) with m terms, the approximate number of calculations would be

$$2m+\sum_{r=1}^{m}\left[10(r+1)+\frac{7}{4}(r+1)(r+3)\right].$$

These considerations imply that computation will be considerably slower for the expansion methods. This suggestion is borne out in table 3. There, to

Table 3

Computation time for comparable accuracy.[a]

N[b]	Expansion (9)	Expansion (10)	Quadrature
7	6.10 (15)[c]	7.51 (15)	0.24 (7)
6	5.21 (14)	5.51 (13)	0.22 (6)
5	3.75 (12)	3.17 (10)	0.19 (5)
4	2.58 (10)	2.06 (8)	0.19 (5)
3	2.05 (9)	1.20 (6)	0.17 (4)
2	0.95 (6)	1.20 (6)	0.13 (3)
1	0.48 (4)	0.60 (4)	0.11 (2)

[a] Times reported are CPU time in seconds on an IBM 3081 Model K with sixteen megabytes of core storage and operating under MVS.
[b] Number of places of accuracy after the decimal point.
[c] Number of terms in expansion, or number of grid points.

minimize 'fixed' time,[8] the same integral was calculated repeatedly (1000 times) for various values of m and n, chosen and displayed to compare cases with approximately the same number of accurately calculated significant digits. The conclusion to be drawn from tables 2 and 3 (and many other unreported trials) is that the method of quadrature is much faster than either expansion method.

5. Concluding remarks

This paper has extended the work of Kiefer (1983) by demonstrating the use of Gaussian quadrature in the computation required for ML estimation of duration models. The proposed technique is general with regard to stochastic assumption regarding the hazard and to type and degree of censoring in the data. The results of tables 2 and 3 suggest that quadrature is the preferred method. The results of table 1 (qualitatively unchanged for various specifications of the explanatory variables, choice of innovations considered, and subsets of observations) indicate that as few as two evaluation points may be used. This means that a large number of exploratory estimations may be performed at relatively low cost. For final analyses the number of grid points should be increased, and comparisons made, but it is unlikely that the cost will be prohibitive.

References

Abramowitz, M. and I. Stegun, 1964, Handbook of mathematical functions with formulas, graphs, and mathematical tables, National Bureau of Standards applied mathematic series no. 55 (U.S. Government Printing Office, Washington, DC.)

Butler, J.S. and R. Moffitt, 1982, A computationally efficient quadrature procedure for the one-factor multinomial probit model, Econometrica 50, 761–764.

Kiefer, N., 1983, An integral occurring in duration models with heterogeneity, Economics Letters 11, 251–256.

Leiderman, P.H., D. Babu, J. Kagia, H.C. Kraemer and G.F. Leiderman, 1973, African infant precocity and some social influences during the first year, Nature 242, 247–249.

Schmee, J. and G.J. Hahn, 1979, A simple method of regression analysis with censored data, Technometrics 21, 417.

Turnbull, B., 1974, Nonparametric estimation of a survivorship function with doubly censored data, Journal of the American Statistical Association 69, 169–173.

Romeo, A.J., J.L. Wagner and R.H. Lee, 1982, Prospective reimbursement and the diffusion of new technologies in hospitals, Mimeo. (The Urban Institute, Washington, DC).

Waldman, D.M. and R.H. Lee, 1983, Estimation in a model of adoption of technological innovation, Mimeo. (University of North Carolina, Chapel Hill, NC).

[8] Computer execution was performed from a 'load' module. Therefore, reported CPU times are free of compilation or link-edit time.