MONASH

16/93

MONASH
UNIVERSITY

AUSTRALIA

MAXIMUM LIKELIHOOD ESTIMATION IN BINARY DATA MODELS USING PANEL DATA UNDER ALTERNATIVE DISTRIBUTIONAL ASSUMPTIONS

Chris Orme and Tim R.L. Fry

Working Paper No. 16/93

October 1993

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0384 6

MAXIMUM LIKELIHOOD ESTIMATION IN BINARY DATA MODELS USING PANEL DATA UNDER ALTERNATIVE DISTRIBUTIONAL ASSUMPTIONS

Chris Orme and Tim R.L. Fry

Working Paper No. 16/93

October 1993

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

Maximum Likelihood Estimation in Binary Data Models using Panel Data under Alternative Distributional Assumptions

by

Chris Orme*
Department of Economics and Related Studies
University of York
York YO1 5DD
United Kingdom

and

Tim R.L. Fry
Department of Econometrics
Monash University
Clayton, Victoria 3168
Australia

Abstract

This note considers a model of (recurrent) univariate binary outcomes which incorporates random individual effects. Given simplifying distributional assumptions, a likelihood can easily be obtained having the attractive feature of being the product of contributions which only involve *sums* and *no numerical integration*. A recent paper by Conaway (1990) considers the same problem but solves it by finding expressions for the probabilities of all the 2^T possible sequences of the T recurrent binary outcomes, some of which will not be observed in a given data set. The approach adopted in this paper derives an expression for the appropriate likelihood given a particular set of data. The likelihood, score vector and hessian matrix can all be written in simple forms which readily permits the use of Newton-Raphson/gradient methods to locate the roots of the score equations. Simulation experiments suggest that convergence is rapid and also provide evidence on the robustness of the model to distributional misspecification.

J.E.L. Classification: C23.

**The initial draft of this paper was written whilst the first author was a Visiting Fellow in The Department of Statistics at The Australian National University, Canberra. The hospitality of the department is gratefully acknowledged. The first author would also like to thank the ESRC for support under Grant R000232681. We are grateful for comments from participants at the Australasian Meeting of the Econometric Society, University of Sydney, 7th - 9th July 1993.*

Address for correspondence: Department of Economics and Related Studies, University of York, Heslington, York YO1 5DD.
Email: cdo2@sargan.york.ac.uk

1. Introduction.

In this paper we derive a tractable likelihood function for the situation in which the econometrician has binary responses (indicating the presence or absence of an event) for a collection of n individuals over T equally spaced discrete time periods (see Hsiao, Chapter 7, (1986)). We assume a static model which includes unobserved random individual effects (or, perhaps, dynamic models in which the relevant pre-sample history can realistically be regarded as truly exogenous) so that the problem of *initial conditions* (see Hsiao (1986, p.169)) will not be an issue. In such a model primary interest focusses on the estimation of the probability that the event of interest occurs (for a particular individual at a given time). Such probabilities could be estimated separately in each time period using simple probit or logit models. The panel nature of the data, however, permits the evaluation of the correlation that may exist between the occurrence, or otherwise, of the event from one time period to the next. Moreover, exploiting the panel will, presumably, lead to more efficient estimation. The most commonly used specification is that of a probit model with random effects (see, for example, LIMDEP version 6 (Greene (1991))).

The distributional assumptions adopted in this paper permit the likelihood to be expressed as a sequence of sums involving no numerical integration. A recent paper by Conaway (1990) has also discussed the form of such a likelihood. Adopting the same statistical framework as ours, Conaway provides a way of calculating all the probabilities corresponding to the 2^T different discrete responses that could arise. The appropriate form for the likelihood is then implied using these probabilities. The derivation in this paper provides a simple expression for the likelihood obtained directly from the observed set of responses. This approach has an advantage when T is even only moderately large since contributions to the likelihood involve easily calculated sums, whereas the method detailed by Conaway would involve the calculation of essentially superfluous probabilities and the inversion of a $(2^T \times 2^T)$ matrix. Moreover, simple expressions for the score vector and hessian matrix are also provided in this paper thus readily enabling standard Newton-Raphson/gradient methods to be employed.

The layout of the paper is as follows: section 2 details the statistical model and concludes by providing a simple expression for the likelihood. Section 3 gives the necessary derivations required to implement numerical optimisation procedures designed to locate maximum likelihood estimates as roots of the score equations. In section 4 we report findings from some simulation experiments.

2. The model.

We begin by postulating a binary panel data model, conditional on random individual specific effects. The observed behaviour of individual i in time period t derives from

$$y^*_{it} = x_{it}'\beta + u_{it} + \varepsilon_i \tag{1}$$

where the observed response is $y_{it} = 1$ if $y^*_{it} > 0$ and zero otherwise and the x_{it}' is a $(1 \times p)$ vector of static covariates, some of which may be time invariant. In particular, we shall assume that it contains a constant term, implying an intercept parameter β_1. The errors terms u_{it} are independently and identically distributed Extreme Value (log-Weibull) variates having cumulative distribution function exp(-exp(-u)), at the point u.

Dropping the individual identifier, for ease of notation, the probability that $y_t = 1$, conditional on ε, is thus $1 - \exp(-\exp(x_t'\beta+\varepsilon))$. This can be expressed as $1 - \exp(-\nu d_t)$, where $d_t = \exp(x_t'\beta)$ and $\nu = \exp(\varepsilon) > 0$. The contribution to the likelihood based on that individual's observed event history, $y' = (y_1,...,y_T)$, but conditional on ν, is written

$$\ell(\nu) = \prod_{t\in A} \exp(-\nu d_t) \prod_{t\notin A} \{1 - \exp(-\nu d_t)\} \tag{2}$$

where $A = \{t; y_t = 1\}$. Expanding the products we obtain

$$\begin{aligned}
\ell(\nu) = {} & \exp\left(-\nu \sum_{k\notin A} d_k\right) - \sum_{t\in A} \exp\left(-\nu\left(d_t + \sum_{k\notin A} d_k\right)\right) \\
& + \sum\sum_{s<t\in A} \exp\left(-\nu\left(d_s + d_t + \sum_{k\notin A} d_k\right)\right) \\
& - \sum\sum\sum_{r<s<t\in A} \exp\left(-\nu\left(d_r + d_s + d_t + \sum_{k\notin A} d_k\right)\right) \\
& + (-1)^z \exp\left(-\nu \sum_{t=1}^{T} d_t\right)
\end{aligned} \tag{3}$$

where $z = \sum_{t=1}^{T} y_t$, (i.e., the total number of 'ones').

In order to get appropriate likelihood contributions we integrate out the random effects term, ν, by taking the expected value of the expression given in (3); i.e., we evaluate $E[\ell(\nu)] = \int_0^\infty \ell(\nu)h(\nu)d\nu$. This requires the adoption of a parametric form for $h(\nu)$, the density of ν. For simplicity we assume that ν is a gamma random variable having unit mean and variance $\eta > 0$. This specification is not uncommon when considering problems of neglected heterogeneity; see, for example, Lancaster (1985), (1990), Crowder (1985), Conaway (1990) and Fry (1988).

When taking the expected value of (3) we simply note that we require the expectation of sums. Further, the expectation of each term in the sum can be obtained from the definition of the moment generating function of a gamma random variable. We thus obtain, marginal with respect to ν, the likelihood contribution as

$$
\begin{aligned}
\ell^* = E[\ell(\nu)] = &\Big(1 + \eta \sum_{k \notin A} d_k\Big)^{-\eta^{-1}} - \sum_{t \in A}\Big(1 + \eta\Big(d_t + \sum_{k \notin A} d_k\Big)\Big)^{-\eta^{-1}} \\
&+ \sum\sum_{s<t \in A}\Big(1 + \eta\Big(d_s + d_t + \sum_{k \notin A} d_k\Big)\Big)^{-\eta^{-1}} \\
&- \sum\sum\sum_{r<s<t \in A}\Big(1 + \eta\Big(d_r + d_s + d_t + \sum_{k \notin A} d_k\Big)\Big)^{-\eta^{-1}} \\
&+ (-1)^z\Big(1 + \eta\sum_{t=1}^{T} d_t\Big)^{-\eta^{-1}} \qquad (4)
\end{aligned}
$$

and the likelihood for the sample of all individual histories is the product of the individual contributions. From (4) we see that a typical contribution only has z+1 'components'. Such a contribution is not too onerous to compute if typically the total number of occurrences, z, of the event of interest is small.

In his paper Conaway notes that the log of this likelihood (for individual i) can be expressed as $\log(\ell^*) = \sum_y \delta_y \log(\pi_y)$, where the δ_y are a sequence of 2^T indicator variables which equal unity if individual i is observed to have event history y' (and zero otherwise) and π_y is the marginal probability of this event history. Conaway then details a method (involving the inversion of a $(2^T \times 2^T)$ matrix) whereby all these 2^T probabilities can be obtained, although in a given application some of these probabilities will not be needed. The resulting expression is then used to construct the likelihood. The approach adopted in this paper does not require the calculation of superfluous probabilities. Moreover, in the next section, we provide expressions for the first and second derivatives of the log-likelihood.

Note that the model described by the likelihood given in (4) obtains when the compound errors $v_t = u_t + \varepsilon$ $(t = 1,...,T)$ have a *multivariate Burr Type II* (see Fry (1988), (1993)) distribution with fixed correlation structure given by

$$\mathrm{corr}(v_t, v_s) = \frac{\psi'(\eta^{-1})}{\psi'(\eta^{-1}) + \psi'(1)}, \qquad t \neq s$$

where $\psi'(.)$ denotes the *trigamma* function. In general, the random vector $\zeta' = (\zeta_1,...,\zeta_q)$ has a multivariate Burr distribution if it has a distribution function given by [1]

$$\mathcal{F}(\zeta') = \left\{ 1 + \eta \sum_{j=1}^{q} \exp(-\zeta_j) \right\}^{-\eta^{-1}}.$$

The marginals of this joint distribution are also Burr. Thus it is easy to express the probability that $y_t = 0$ as

$$\Pr[y_t = 0] = \Pr[v_t \leq -\beta' x_t] = [1 + \eta \exp(\beta' x_t)]^{-\eta^{-1}}.$$

Using the joint distribution function, given above, we can express the individual likelihood contributions, (4), in a more compact manner. Firstly, for ease of notation we write $\alpha = \eta^{-1}$ and re-define the intercept term in the regression to be $\beta_1 - log(\alpha)$. Now let $A_j = \{t_{(1)}, t_{(2)}, ..., t_{(j)}\}$ be any

1 The multivariate density function follows immediately and is uni-modal at $\zeta' = 0'$. This density is different to that given by Fry (1989) since it results from slightly different mixing arguments. In the limit, as $\eta \to 0$ we obtain the multivariate distribution function for a sequence of independent extreme value random variables. The mixing arguments of Fry (1988), on the other hand, do not readily provide such a non-degenerate limiting case.

subset of j integers from the set $A = \{t;\ y_t = 1\}$ [2]; clearly, there are ${}_zC_j = \binom{z}{j}$ such combinations of these j integers. Define the following:

$$w_0 = \sum_{k \notin A} d_k \quad \text{and} \quad w_j = \sum_{r \in A_j} d_r,\ j = 1, \ldots, z,$$

observing that there are ${}_zC_j$ ways of obtaining w_j.[3] For example, if A contains the elements {1,2,3,4}, then w_2 can be expressed as $(d_1 + d_2)$, $(d_1 + d_3)$, $(d_1 + d_4)$, $(d_2 + d_3)$, $(d_2 + d_4)$ or $(d_3 + d_4)$. With these preliminaries it will be useful to define a sequence of (Burr distribution) functions as

$$F_0^{\alpha} = [1 + w_0]^{-\alpha} \quad \text{and} \quad F_j^{\alpha} = [1 + w_0 + w_j]^{-\alpha},\ j = 1, \ldots, z,$$

with $G_j = \sum_{A_j \subset A} F_j^{\alpha},\ j = 0, \ldots, z,$ where, the sum is taken over all possible choices for A_j with $G_0 = F_0^{\alpha}$ and $G_z = F_z^{\alpha} = [1 + \sum_{t=1}^{T} d_t]^{-\alpha}$. Again, using the previous simple example for illustrative purposes, we see that

$$\begin{aligned} G_2 = {} & [1 + w_0 + (d_1 + d_2)]^{-\alpha} + [1 + w_0 + (d_1 + d_3)]^{-\alpha} \\ & + [1 + w_0 + (d_1 + d_4)]^{-\alpha} + [1 + w_0 + (d_2 + d_3)]^{-\alpha} \\ & + [1 + w_0 + (d_2 + d_4)]^{-\alpha} + [1 + w_0 + (d_3 + d_4)]^{-\alpha}. \end{aligned}$$

With this simplifying notation, the likelihood for individual i is thus

$$\ell = \sum_{j=0}^{z} (-1)^j G_j \tag{5}$$

giving the log-likelihood, given a sample of n individual event histories, as $\mathcal{L}(\theta) = \sum_{i=1}^{n} log(\ell_i)$, where $\theta = (\beta', \alpha)'$ is the $(p+1 \times 1)$ vector of unknown parameters.

Having derived a simple expression for the likelihood, which only involves relatively straightforward sums, we can now quite easily obtain first and second order partial derivatives of $\mathcal{L}(\theta)$. These can be used in a numerical optimisation procedure in order to locate the roots of the score

2 We have the identity, $A_z \equiv A$.

3 Observe that $w_0 + w_z = \sum_{t=1}^{T} d_t$.

equations $\partial\mathcal{L}(\hat{\theta})/\partial\theta_l = 0$, $l = 1,\dots,p+1$, where $\hat{\theta}$ denotes the maximum likelihood estimate of θ. The required derivations are given in the next section.

3. Model estimation.

Given the log-likelihood $\mathcal{L}(\theta)$, as defined above, we have that $\partial\mathcal{L}(\theta)/\partial\theta' = \sum_{i=1}^{n}\partial log(\ell_i^*)/\partial\theta' = \sum_{i=1}^{n}\ell_i^{*-1}\cdot\partial\ell_i^*/\partial\theta'$; and as ℓ_i^* is a sum of terms itself these derivatives are clearly manageable. In particular we can write, dropping the observational subscript,

$$\partial\ell^*/\partial\theta = \sum_{j=0}^{z}(-1)^j\partial G_j/\partial\theta = \sum_{j=0}^{z}(-1)^j\sum_{A_j\subset A}\partial F_j^{\alpha}/\partial\theta.$$

To proceed, we introduce the following notation (which is similar in spirit to that used at the end of section 2),

$$m_0 = \sum_{k\notin A} d_k x_k \; ; \quad m_j = \sum_{r\in A_j} d_r x_r,\ j = 1, \dots, z,\ (p \times 1);$$

$$M_0 = \sum_{k\notin A} d_k x_k x_k' \; ; \quad M_j = \sum_{r\in A_j} d_r x_r x_r',\ j = 1, \dots, z,\ (p \times p).$$

Now, first and second-order partial derivatives of F_j^{α} are given by

$$\partial F_j^{\alpha}/\partial\theta = \begin{pmatrix}\partial F_j^{\alpha}/\partial\beta \\ \partial F_j^{\alpha}/\partial\alpha\end{pmatrix} = \begin{pmatrix}-\ \alpha F_j^{\alpha+1}(m_0 + m_j) \\ -\ F_j^{\alpha} log(1 + w_0 + w_j)\end{pmatrix},$$

$$\partial^2 F_j^{\alpha}/\partial\theta\partial\theta' = \left(\begin{array}{c:c}\alpha(\alpha+1)F_j^{\alpha+2}(m_0 + m_j)(m_0 + m_j)' & -\ F_j^{\alpha+1}(1 + logF_j^{\alpha})(m_0 + m_j) \\ -\ \alpha F_j^{\alpha+1}(M_0 + M_j) & \\ \cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots \\ -\ F_j^{\alpha+1}(1 + logF_j^{\alpha})(m_0 + m_j)' & F_j^{\alpha}[log(1 + w_0 + w_j)]^2\end{array}\right).$$

The $(p+1 \times 1)$ score vector can thus be written as (summing over individuals)

$$\partial\mathcal{L}(\theta)/\partial\theta = \sum_{i=1}^{n}\partial log\ell_i^*/\partial\theta = \sum_{i=1}^{n}\{\ell_i^*\}^{-1}\left\{\sum_{j=0}^{z}(-1)^j\sum_{A_j\subset A}\partial F_j^{\alpha}/\partial\theta\right\},$$

and the (p+1 $\times$ p+1) hessian as

$$\partial^2\mathcal{L}(\theta)/\partial\theta\partial\theta' = \sum_{i=1}^{n}\left(\{\ell_i^*\}^{-1}\left\{\sum_{j=0}^{z}(-1)^j \sum_{A_j\subset A} \partial F_j^{\alpha}/\partial\theta\partial\theta'\right\} - \frac{\partial log\ell_i^*}{\partial\theta}\frac{\partial log\ell_i^*}{\partial\theta'}\right.$$

These expressions can then be used in a standard Newton-Raphson type iterative procedure to locate the maximum likelihood estimate. Experimentation with simulated data suggests that convergence is very rapid.

4. Simulation evidence.

The simulation evidence presented in this section arises from three experiments. Firstly we assess the performance of the 'Burr' maximum likelihood estimates when data are generated according to the distributional assumptions of section 2; i.e., the compound errors v_t are multivariate Burr with parameter $\eta > 0$, so that ℓ^* defined at (3) provides the correct likelihood contributions. Secondly, the properties of these maximum likelihood estimates are investigated when the data are actually generated according to a *probit* specification; i.e., the unobservables, u_{it} and ε_i in equation (1) are both normally distributed (with the variance of u_{it} normalised at unity) so that the compound errors, v_t, are normally distributed mean zero, variance $(1-\rho)^{-1}$ and correlation structure $corr(v_t,v_s) = \rho$. Finally, we use a data generation process where both u_{it} and ε_i are chi-square random variables. For all three experiments we assess the ability of the estimated model to correctly predict the probabilities associated with the particular event in a given time period; thus, for individuals observed repeatedly over T time periods, we wish to estimate T probabilities, $Pr[y_t = 0]$, $t = 1,...,T$ (or $Pr[y_t = 1]$). We will therefore not be concerned directly with estimation of parameters.

Data are generated according to the specification given at equation (1) in which four regressors, x_{itl}, are employed:

$x_{it1} = 1$, for all i and t;

$x_{it2} = U(2,10)$, a *uniform* random variable held fixed over t;

$x_{it3} = 1$ if $x^*_{it3} > 0$ and zero otherwise, where x^*_{it3} is generated according to $x^*_{it3} = \delta(x_{it2}-6) + e_i$, $e_i \sim N(0,1)$, and δ is chosen so that the correlation between x^*_{it3} and x_{it2} is 1/2. The dummy variable x_{it3} is also held fixed over t;

$x_{it4} = 0.1t + 0.5x_{i,t-1,4} + U(-0.5,0.5)$, which varies over i and t.

These regressors were all held fixed in repeated sampling. In all experiments 1000 replications of sample data were employed and the values of the regression coefficients in equation (1), β_l, $l = 1,...,4$, were -3, 0.3, 1 and 1 respectively. Further, for the results reported here n = 250 and T = 4.

Probabilities of interest will be denoted $\pi_t = \Pr[y_t = 0]$, $t = 1,...,4$, and are calculated as follows. Let $c_t = -\beta' x_t$, $t = 1, ...,4$ where x_t is a vector containing given values of the regressors (common to all individuals) at time t. Then the 'true' probabilities can be obtained using the marginal distribution functions for v_t, $t = 1, ...,4$ as defined by the data generation process (either Burr, Normal, or Chi-squared):

$$F_t = \Pr[v_t \leq c_t],\ t = 1, ...,4.$$

Using the Burr model we can obtain estimates of these probabilities by replacing the true parameter values in the above with the obtained maximum likelihood estimates.[4] In the experiments reported below the estimated probabilities (and true probabilities) are calculated when the regressor variable, for time period t, is evaluated at its sample mean over individuals, except x_{t3} which takes the value 1.[5] The comparison with the true probabilities is then made by taking the means of these estimated probabilities over the 1000 replications of sample data. The results are summarised in Table 1.

Experiment A: The Burr Model

Here we report the results over 1000 replications of sample data when n = 250, T = 4, v_t, $t = 1, ...,4$ distributed multivariate Burr with $\eta = 2$; the implied correlation between v_t and v_s is thus 0.75. For maximum likelihood estimation we use the parameterisation adopted in section 3 (i.e., write $\alpha = \eta^{-1} = 0.5$ and re-define the intercept to be $\beta_1 - log(\alpha)) = -2.7$. Although a full set of simulation results are not reported here (but are available upon request) we note that, over the 1000 replications, the mean of the MLE for α

---

4 Note that when evaluating the true probabilities, the functional forms for F_t are given by the assumptions of data generation process. When constructing the implied estimated probabilities, the functional forms for F_t are given by the assumptions underlying the estimated model.

5 These values of x_{tl}, $l=1,2,3$, will be the same for all time periods since the first three regressors are time invariant. The sample mean of x_{t4}, however, changes with t.

was 0.51 compared to the true value of 0.5. When estimating the value of the β parameters in (1), the means of the maximum likelihood estimates were -3.01, 0.30, 1.02 and 1.01, very close to the true parameter values. The means of the estimated standard errors (using the estimated hessian) were also very close to the observed standard deviation of the maximum likelihood estimates. Individual *t-ratios* are constructed of the hypotheses β_2=0.3, β_3=1.0, β_4=1.0 and α=0.5; over the 1000 replications the observed rejection frequencies (at the nominal 5% level) are 4.8%, 4.0%, 5.5% and 4.8%, respectively, which appear reasonable. For this experiment, and in others not reported here, the optimisation routine exhibited rapid convergence.

Experiment B: The Probit Model

In this experiment 1000 replications of sample data were taken with n = 250, T = 4, and where v_t, t = 1, ...,4 is multivariate Normal with zero means, variances equal to 2 and a correlation of ρ = 0.5. Using this data generation process we again obtain 'Burr' maximum likelihood estimates and, using these, estimate (under the Burr assumption) the probabilities discussed above. The fact that v_t, t = 1, ...,4 is actually multivariate Normal means that the corresponding true probabilities can easily be constructed as $\pi_t = \Phi(c_t/\sqrt{2})$, where $\Phi(.)$ denotes the standard normal *cdf.*

Experiment C: Chi-square errors

In this experiment 1000 replications of sample data were taken with n = 250, T = 4, where u_t was a standardised χ^2_4 random variable distributed independently of ε which was a centred χ^2_2 variate (i.e., having zero mean and variance equal to 4) and scaled to have variance 0.5. Thus, in an obvious notation, $v_t = (\chi^2_6 - 6)/\sqrt{8}$. 'Burr' maximum likelihood estimates are obtained and used to obtained estimated probabilities, π_t. The true probabilities are $\pi_t = \Pr[\chi^2_6 \leq c_t\sqrt{8} + 6]$.

The mean of the estimated probabilities, over the 1000 replications, and the true probabilities are presented below in Table 1. Under each experiment label the true probabilities are given in the left hand column and the corresponding mean of the estimated probabilities are given in the right hand column. All estimated results in Table 1 are based on the assumption of a Burr likelihood as defined by equations (4) and (5). The results reported under *Experiment A* simply confirm that maximum likelihood estimation does well at estimating the appropriate probabilities under a correctly specified Burr model.

The encouraging result, under *Experiment B* and *C*, is that the (Burr) estimated probabilities are still close to the true probabilities even when the the compound errors ($v_t = \varepsilon + u_t$) are not, themselves, multivariate Burr. Although the divergence between true and estimated is greater than under Experiment A, the agreement is remarkably close.

Table 1: Burr Results						
	A: $v_t \sim$ Burr $\eta = 2$		*B:* $v_t \sim$ Normal $\rho = 0.5$		*C:* Chi-square $v_t = (\chi^2_6 - 6)/\sqrt{8}$	
	True	Estimated	True	Estimated	True	Estimated
π_1	0.630	0.630	0.577	0.589	0.658	0.649
π_2	0.592	0.590	0.521	0.531	0.600	0.583
π_3	0.559	0.558	0.474	0.483	0.547	0.525
π_4	0.515	0.514	0.409	0.417	0.466	0.442

Notes: In Experiment C, u_t is standardised to have zero mean and unit variance, ε has zero mean and variance equal to 0.5.

It is also worth noting that, under Experiment B, the mean values of the MLE's for β_2, β_3 and β_4 are (respectively) 1.17, 1.16 and 1.18 times the true values of these coefficients. This suggests that there is simply a re-scaling effect taking place (rather like the difference between logit and probit) but this does not adversely affect the ability of the model to estimate the required probabilities.[6] In order to investigate this a bit further we can look at the true marginal distribution function of v_t, which is $\Phi(v/\sqrt{2})$ and compare it with the fitted marginal distribution obtained from the Burr estimates. The fitted marginal distribution function is constructed as

$$F(v) = [1 + \frac{1}{a} \exp(-v/\tau)]^{-a}$$

6 A similar re-scaling effect is not observed for the estimated intercept term. This is probably due to the fact whilst the bivariate normal distribution has a zero mean vector, this is not necessarily true of the bivariate Burr. On the assumption that the estimation routine will endeavour to provide a Burr distribution which approximates the true normal distribution then there may be a mean adjustment, through the estimated $\alpha = \eta^{-1}$ parameter, which will be absorbed into the estimated intercept.

where a = 0.9575 is the mean of the Burr estimates for α and $\tau = 1/1.17$ is the observed scaling factor. The two distribution functions are graphed in Figure 1. There seems to be very close agreement between the two distribution functions, except in the extreme tails. This provides some further evidence on the ability of the Burr model to adequately mimic a probit specification.

---insert figure 1 here

In the case of Experiment C, a similar rescaling effect is not so clearly apparent. The actual mean values of the MLE's for β_j, j = 2,3,4 are 0.36, 1.17 and 1.26, in this case. Nonetheless, the estimated probabilities reported in Table 1, under this experiment, are close to (but slightly underestimate) the 'true' probabilities.

Finally, in order to a compare the performance of Burr against probit formulations, estimated probabilities were obtained using a probit specification for experiments A and C described above. Exactly the same set of (fixed in repeated samples) regressors were employed, but simulation results were obtained using LIMDEP (Greene, 1991) using 100 samples of replicated data. Under Experiment B, 100 replications gave estimated probabilities which were very close to their true values. The results for Experiments A and C are reported in Table 2.

Table 2: Probit Results using LIMDEP

	A: v_t~ Burr $\eta = 2$		*C:* Chi-square $v_t = (\chi^2_6 - 6)/\sqrt{8}$	
	True	Estimated	True	Estimated
π_1	0.630	0.625	0.658	0.623
π_2	0.592	0.590	0.600	0.559
π_3	0.559	0.560	0.547	0.504
π_4	0.515	0.518	0.466	0.428

The probit specification does well in Experiment A when the data are generated according to a Burr specification and again the estimated slope coefficients in the regression function are rescaled by a factor of about 0.79. However, when the data are generated according to the chi-squared specification then the probit model underestimates the true probabilities of interest and, significantly, does not perform as well as the Burr model. This must be due, in part, to the flexibility of the Burr formulation.[7]

5. Conclusion.

In this paper we have provided computationally simple expressions which facilitate maximum likelihood estimation of parameters in an econometric model of recurrent binary outcomes, incorporating individual random effects. The simplifying stochastic assumptions leads to an attractive expression for the log-likelihood and associated partial derivatives which only involve elementary sums. Thus numerical integration, which characterises alternative model specifications (such as the multivariate probit), has been replaced by arithmetically tractable sums.

Although this paper refers to 'event histories' the methodology, and algebra, used here could equally be well applied to multivariate choice models in which an individual can make any number of choices (not just one) from a set of T alternatives (see Fry (1988) for a discussion of a simple tri-variate choice model).

Simulation experiments revealed rapid convergence using a simple Newton-Raphson optimisation routine, even when the number of repeated spells is as large as 5. The results presented in section 4 also show that the (Burr) estimation of relevant probabilities can remain accurate when the data are generated under alternative distributional assumptions. In the simulations reported here, the Burr specification adequately modelled normality and proved to be more flexible than the probit specification under alternative chi-squared distributional assumptions. Thus the Burr model suggests itself as a relatively simple alternative to the more commonly adopted (and restrictive) probit model.

7 See also Fry and Orme (1993) for an application of the Burr model in Tobit-type (censored) linear regressions having normal and chi-squared errors.

References

Conaway, M.R. (1990), "A Random Effects Model for Binary Data", *Biometrics*, **46**, 317-328.

Crowder, M. (1985), "A Distributional Model for Repeated Failure Time Measurements", *Journal of the Royal Statistical Society (B)*, **47**, 447-452.

Fry, T.R.L. (1988), *On the Use of Some Burr Family Distributions in Econometrics*. Unpublished Ph.D. Thesis, University of Manchester.

Fry, T.R.L. (1993), 'Univariate and Multivariate Burr Distributions: A Survey', *The Pakistan Journal of Statistics*, **9**, 1-24.

Fry, T.R.L. and C. Orme (1993), 'A Generalized Logistic Tobit Model', *mimeo., Department of Econometrics, Monash University.*

Greene, W.H. (1991), *Limdep, Version 6*, Econometric Software, New York.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Lancaster, T. (1985), 'Generalised Residuals and Heterogeneous Duration Models: The Weibull Case', *Journal of Econometrics*, **28**, 155-169.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Figure 1