Economics Letters 19 (1985) 31-34
North-Holland

ESTIMATING TRANSITION PROBABILITIES FROM PANEL DATA *

Marcel G. DAGENAIS
University of Montreal, Montreal, Que., Canada H3C 3J7

Alice NAKAMURA and Masao NAKAMURA
University of Alberta, Edmonton, Alb., Canada T6G 2E1

Received 12 February 1985

This paper discusses the maximum likelihood estimation of conditional probabilities for dichotomous events, when using panel data and controlling for individual characteristics.

The modeling and estimation of the probability of joint dichotomous events is an established topic in the economics literature [see, for instance, Amemiya (1981) and Poirier (1980)]. In other disciplinary areas such as marketing, management science and demography, however, conditional rather than joint probabilities are typically estimated [see, for instance, Massey, Montgomery and Morrison (1970), Keyfitz (1968), Nakamura (1973), and Nakamura and Nakamura (1977,1978)]. The conditional probabilities which often represent transition probabilities are usually estimated without controlling for heterogeneity due to observable variables and without attention to possible biases resulting from sample selection. For instance, in discussing a transition matrix for the selection of brand 1 in period t depending on whether or not brand 1 was selected in $t-1$, Massey, Montgomery and Morrison (1970, p. 84) write: 'It is obvious that the transition matrix can tell us only the combined effect of *both* heterogeneity of the population *and* the effect of the past purchase.'

In certain modeling and forecasting problems in economics, conditional rather than joint probabilities are also required. In a microsimulation model of the household sector, the characteristics of the individuals in an initial population are updated on a year-by-year basis [see, for instance, Orcutt et al. (1961), Yett et al. (1975), Orcutt, Caldwell and Wertheimer (1976), and Nakamura and Nakamura (1978)]. It has been established that the work/non-work decision of an individual in year $t-1$ embeds information about persistent unobservables that have important impacts on the individual's probability of work in year t [see Nakamura and Nakamura (1985,n.d.)]. Thus the required probability of work in a microsimulation model is the conditional probability of work in year t given the observed (or simulated) work status of the individual in year $t-1$. The purpose of this note is to show that, building on the literature on the estimation of joint probabilities for dichotomous events, conditional probabilities for events of this sort can easily be estimated controlling for heterogeneity

* This note is the outgrowth of discussions initiated at the Canadian Econometric Study Group 1984 meetings, at Queen's University. The authors benefitted especially from the comments of Paul Ruud, Dale Poirier and Jean-Marie Dufour. The authors are solely responsible for any remaining errors.

0165-1765/85/$3.30

due to observable factors and potential selection bias problems.

Suppose each individual is faced with choosing between A and B in both period 1 and period 2. The conditional probabilities of interest are $P(A_2 | A_1)$, $P(A_2 | B_1)$, $P(B_2 | A_1)$ and $P(B_2 | B_1)$, where $P(A_2 | A_1)$, for example, stands for the conditional probability of choosing alternative A in period 2 given that alternative A was chosen in period 1. Suppose also that the values of these conditional probabilities differ systematically from individual to individual due to the impacts of observable individual characteristics. If A stands for work and B for non-work, for instance, the values of the relevant conditional probabilities might be expected to differ systematically for women depending on factors such as age, education and child status.

Adopting a stochastic utility approach, suppose that, say, alternatives A_1 and A_2 will be chosen consecutively in periods 1 and 2 by the jth individual if

$$U_j(A_1) - U_j(B_1) = X_{1j}\beta + u_{1j} > 0 \quad \text{and} \quad U_j(A_2) - U_j(B_2) = X_{2j}\beta + u_{2j} > 0, \tag{1),(2}$$

where, for example, $U_j(A_1)$ designates the utility for individual j of choosing alternative A in period 1, X_{1j} is the vector of characteristics of individual j (such as age, education, wealth, etc.) prevailing in period 1, β is the vector of parameters, and u_{1j} and u_{2j} are jointly normally distributed residual errors with mean vector 0 and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Note that in this model, the elements of the parameter vector $(\beta', \sigma_{11}, \sigma_{22}, \sigma_{12})$ are specified only up to some arbitrary scalar. One can therefore set a priori $\sigma_{11} = 1$. Clearly, in the majority of cases, one would expect to have $\sigma_{12} > 0$. In many cases, it would also be reasonable to assume $\sigma_{22} = \sigma_{11}$.

Any conditional probability can be expressed in terms of joint and marginal probabilities. Thus from (1) and (2), we can express the conditional probabilities of interest for individual j, such as $P_j(A_2 | A_1)$, as follows:

$$P_j(A_2 | A_1) = P_j(A_2, A_1)/P_j(A_1) = \int_{-X_{2j}\beta}^{\infty} \int_{-X_{1j}\beta}^{\infty} f(u_{1j}, u_{2j}) \mathrm{d}u_{1j} \mathrm{d}u_{2j} \bigg/ \int_{-X_{1j}\beta}^{\infty} f(u_{1j}) \mathrm{d}u_{1j}, \tag{3}$$

where $f(u_{1j}, u_{2j})$ is the joint normal density function of u_{1j} and u_{2j} mentioned above and $f(u_{1j})$ is the corresponding marginal normal density of u_{1j} with mean zero and variance σ_{11}. Similarly, we have

$$P_j(A_2 | B_1) = P_j(A_2, B_1)/P_j(B_1) = \int_{-X_{2j}\beta}^{\infty} \int_{-\infty}^{-X_{1j}\beta} f(u_{1j}, u_{2j}) \mathrm{d}u_{1j} \mathrm{d}u_{2j} \bigg/ \int_{-\infty}^{-X_{1j}\beta} f(u_{1j}) \mathrm{d}u_{1j}, \tag{4}$$

$$P_j(B_2 | A_1) = P_j(B_2, A_1)/P_j(A_1) = \int_{-\infty}^{-X_{2j}\beta} \int_{-X_{1j}\beta}^{\infty} f(u_{1j}, u_{2j}) \mathrm{d}u_{1j} \mathrm{d}u_{2j} \bigg/ \int_{-X_{1j}\beta}^{\infty} f(u_{1j}) \mathrm{d}u_{1j}, \tag{5}$$

$$P_j(B_2 | B_1) = P_j(B_2, B_1)/P_j(B_1) = \int_{-\infty}^{-X_{2j}\beta} \int_{-\infty}^{-X_{1j}\beta} f(u_{1j}, u_{2j}) \mathrm{d}u_{1j} \mathrm{d}u_{2j} \bigg/ \int_{-\infty}^{-X_{1j}\beta} f(u_{1j}) \mathrm{d}u_{1j}. \tag{6}$$

If a random cross-sectional sample of data for N individuals were available for either period 1 or 2, we could obtain maximum likelihood (ML) estimates of β/σ_{11} or of β/σ_{22} respectively, using standard probit analysis. This would allow us to calculate estimates for $P_j(A_1)$, $P_j(B_1)$, $P_j(A_2)$ and

$P_j(B_2)$, but not for the conditional probabilities of interest. Notice also that assuming $\sigma_{22} = \sigma_{11}$, we cannot obtain ML estimates of β/σ_{11}, using pooled data with a standard probit routine, even if data are available for both periods 1 and 2, because of the autocorrelation of the residual errors.

We might think of obtaining estimates of $P_j(A_2 | A_1)$ and $P_j(B_2 | A_1)$ using data for period 2 for the censored sample of those who chose A in period 1. Likewise we might think of obtaining estimates of $P_j(A_2 | B_1)$ and $P_j(B_2 | B_1)$ using data for period 2 for the censored sample of those who chose B in period 1. The residual errors obey truncated normal distributions in these censored samples, however. Thus standard probit analysis cannot be applied.

However, if we have data over periods 1 and 2 for a random sample of N individuals, ML estimates of these probabilities can be calculated from ML estimates of the unknown elements of β and Σ. Such parameter estimates can be obtained by maximizing the likelihood function,

$$L = \prod_{j \in S_1} P_j(A_2, A_1) \prod_{j \in S_2} P_j(A_2, B_1) \prod_{j \in S_3} P_j(B_2, A_1) \prod_{j \in S_4} P_j(B_2, B_1), \tag{7}$$

where S_1 stands for the subset of individuals who chose alternative A in period 2 and alternative A in period 1 and where S_2, S_3 and S_4 are also defined appropriately.

Bivariate normal integrals can be computed efficiently using Divgi's (1979) algorithm. The likelihood function can be maximized using the Fletcher–Powell (1963) algorithm. This algorithm requires the computation of the first derivatives of L with respect to θ, where $\theta = (\beta', \sigma_{22}, \sigma_{12})'$. An estimate of the asymptotic covariance matrix ($\hat{C}$) of the parameter estimates $\hat{\theta}$, can be obtained by using the approach suggested by Berndt et al. (1974).

The conditional probabilities expressed in eqs. (3)–(6) can then be estimated by replacing the unknown elements of θ by those of $\hat{\theta}$; and the asymptotic covariance matrix of these probability estimates can be obtained, for given values of X_{1j} and X_{2j}, by using the Goldberger–Nagar–Odeh (1961) formula,

$$\hat{V}_j = \left\{ \left[\partial P_j/\partial\theta \right]' \hat{C} \left[\partial P_j/\partial\theta \right] \right\}_{\theta=\hat{\theta}}, \tag{8}$$

where $\hat{V}_j$ is the covariance estimate in question, and

$$\partial P_j/\partial\theta = \left[\partial P_j(A_2 | A_1)/\partial\theta, \partial P_j(A_2 | B_1)/\partial\theta, \partial P_j(B_2 | A_1)/\partial\theta, \partial P_j(B_2 | B_1)/\partial\theta \right]. \tag{9}$$

The above approach can be extended readily to the estimation of conditional probabilities other than transition probabilities. For example, one could consider the probability that

- a women who worked (or did not work) last year will have (or will not have) a baby this year,
- a women who was (or was not) divorced last year will (or will not) work this year, and
- a person who did (or did not) buy a ticket for the football season last summer will (or will not) buy a season ticket for the opera this fall.

References

Amemiya, T., 1981, Qualitative response models: A survey, Journal of Econometric Literature 19, 1483–1536.

Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman, 1974, Estimation and inference in nonlinear structural models, Annals of Economic and Social Measurement 3, 653–665.

Divgi, D.R., 1979, Calculation of univariate and bivariate normal probability functions, Annals of Statistics 7, no. 4, 903–910.

Fletcher, R. and M.J.D. Powell, 1963, A rapidly convergent descent method for minimization, The Computer Journal 6, 163–168.

Goldberger, A.S., A.L. Nagar and H.S. Odeh, 1961, The covariance matrices of reduced-form coefficients and of forecasts for a structural econometric model, Econometrica 29, 556–573.

Keyfitz, N., 1968, Introduction to the mathematics of population (Addison-Wesley, Reading, MA).

Massey, W.F., D.B. Montgomery and D.G. Morrison, 1970, Stochastic models of buying behavior (The M.I.T. Press, Cambridge, MA).

Nakamura, M., 1973, Some programming problems in population projections, Operations Research 21, 1048–1062.

Nakamura, A. and M. Nakamura, 1977, A Markov analysis of per capital state and local police expenditures and the allocation problem of federal aid, Operational Research Quarterly 28, 293–304.

Nakamura, A. and M. Nakamura, 1978, On microanalytic simulation and its applications in population projection, Journal of the Operational Research Society 29, 349–360.

Nakamura, A. and M. Nakamura, 1985, The second paycheck: An analysis of the employment and earnings of wives compared with unmarried women and men (Academic Press, New York).

Nakamura, A. and M. Nakamura, n.d., Dynamic models of the labor force behavior of married women which can be estimated using limited amounts of past information, Journal of Econometrics, forthcoming.

Orcutt, G.H., S. Caldwell and R. Wertheimer II, 1976, Policy exploration through microanalytic simulation (The Urban Institute, Washington, DC).

Orcutt, G.H., M. Greenberger, J. Korbel and A. Rivlin, 1961, Microanalysis of socioeconomic systems: A simulation study (Harper and Row, New York).

Poirier, D.J., 1980, Partial observability in bivariate probit models, Journal of Econometrics 12, 209–217.

Yett, D.E. et al., 1975, A microeconomic model of the health care system in the United States, Annals of Economic and Social Measurement 4, 407–433.