



A Random-Effects Logit Model for Panel Data

Wolf, D.A.

IIASA Working Paper

WP-87-104

October 1987



Wolf, D.A. (1987) A Random-Effects Logit Model for Panel Data. IIASA Working Paper. WP-87-104 Copyright © 1987 by the author(s). <http://pure.iiasa.ac.at/2948/>

Working Papers on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work. All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting repository@iiasa.ac.at

WORKING PAPER

A RANDOM-EFFECTS LOGIT
MODEL FOR PANEL DATA

Douglas Wolf

October 1987
WP-87-104



**A RANDOM-EFFECTS LOGIT
MODEL FOR PANEL DATA**

Douglas Wolf

October 1987
WP-87-104

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Heterogeneity in population data, a phenomenon which has been the subject of investigation in IIASA's Population Program for several years, is manifested in numerous ways. In the area of econometrics, heterogeneity is alternatively labelled unobserved or unmeasured variables. Unmeasured variables cause statistical problems when empirical data include repeat observations on the behavior of individual units of analysis, such as people, households, or firms. A classic situation is "time-series cross-section" data, in which a sample of units is observed at regular intervals over time. Among the approaches to such data is what is called a "random effects" model, in which a parametric distribution is postulated for the unmeasured variable. This paper develops a random-effects model for a particular kind of empirical situation, one in which the endogenous variable of interest is binary. The logit functional form is adopted, and a binomial distribution is proposed for the unobserved heterogeneity. Maximum-likelihood estimators for the parameters of the model are derived, and the performance of the estimators is investigated in a series of Monte-Carlo experiments.

ABSTRACT

A random-effects panel logit model is proposed, in which the unmeasured attributes of an individual are represented by a discrete-valued random variable, the distribution of which is binomial with a known number of support points. The maximum-likelihood estimator of the unknown parameters of the model are derived, and the performance of the ML estimators is investigated in a series of Monte-Carlo experiments. Several further extensions of the framework are also suggested, including application to discrete event-history data.

A RANDOM-EFFECTS LOGIT MODEL FOR PANEL DATA

Douglas Wolf

In his recent survey article "Limited Dependent Variable Models Using Panel Data" Maddala (1987) discusses alternative statistical approaches to data consisting of a sequence of individual observations on discrete outcomes. An important problem with which the analyst of such data must deal is the possibility of persistent but unmeasured relevant attributes of the decisionmaker. The effect of such unmeasured attributes is to introduce correlations among the disturbances of the equations governing each choice in an individual's sequence. Several statistical models for cross-sectional data – i.e. data for a single outcome for each individual – have been generalized to deal with repeated observations. These generalizations can be grouped into two broad categories, "fixed-effects" and "random-effects" models. Maddala discusses several advantages of the random-effects approach (see Maddala, 1987: 309), yet he dismisses the possibility of a random-effects logit model, on the grounds that it must be based upon the multivariate logistic distribution: the multivariate logistic distribution has the undesirable property that all correlations between pairs of disturbances are constrained to equal one-half.

It is the objective of this note to propose an alternative generalization of the logit model: a random-effects model in which the disturbances are not constrained to have any particular value (they are, however, constrained to be nonnegative). Specifically, it is assumed that the unmeasured individual attributes, or heterogeneity, can be represented by a discrete distribution, the binomial with a known number of support points. In the special case of two support points, heterogeneity consists of an unmeasured dummy variable. The model proposed generalizes the work of Rosenbaum and Rubin (1983), who consider the effect of an unmeasured dummy variable in a single-period binary-logit equation. The

parameters of Rosenbaum and Rubin's model are not all identified; rather, their model serves as a framework for determining the sensitivity of the coefficient on measured covariates to a range of plausible values of the unidentified parameters.¹ In the model discussed below, identification is achieved through the use of repeated observations on an individual's discrete outcome.

In the following section, the model is specified, and maximum-likelihood estimators of its parameters are derived. Only binary outcomes are considered, but the model could easily be extended to deal with multivariate outcomes. This is followed by results from several sampling experiments designed to explore the performance of the ML estimators when the data are generated by the assumed model. A concluding section discusses several possible extensions of the basic approach.

THE MODEL

We begin with Maddala's equation,

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}, \quad i=1, \dots, N; t=1, \dots, T, \quad (1)$$

in which y_{it} is a latent variable, α_i is an individual-specific and time-invariant effect, x_{it} is a (column) vector of exogenous variables at time t (including a constant), β' is a (row) vector of unknown regression coefficients, and u_{it} is a disturbance term assumed uncorrelated across individuals and over time for an individual. The observable outcome d_{it} is an indicator of the sign of y_{it} . We assume that the distribution function for the u_{it} is logistic, eg. $F(u) = [1 + e^{-u}]^{-1}$, so that conditional on α_i the observed dependent variable d_{it} equals one with probability $\exp(\alpha_i + \beta' x_{it}) [1 + \exp(\alpha_i + \beta' x_{it})]^{-1}$ and equals zero with probability $[1 + \exp(\alpha_i + \beta' x_{it})]^{-1}$.

¹ It should be noted that Rosenbaum and Rubin permit the unmeasured variable to be correlated with the measured covariate; this correlation, in fact, motivates their analysis. The model considered here, however, assumes independence of the measured and unmeasured covariates. An interesting application of Rosenbaum and Rubin's approach can be found in Montgomery et al. (1986). Montgomery et al. also generalize the approach, and in one instance achieve identification of the model through what is, in effect, a discrete "multiple indicator" specification. Another, quite different approach to the problem of serial dependencies in a panel logit model can be found in Zeger et al. (1985).

Now, suppose that $\alpha_i = \delta m_i$, where m is a binomially-distributed random variable with parameters $M, s; 0 < s < 1, M$ a positive integer. In other words, $pr(m=j) = \pi_j = \binom{M}{j} s^j (1-s)^{M-j}, j=0, \dots, M$. We assume $cov[x_{it}, m_i] = 0$. The expected value of m is Ms . In the simplest case, $M=1$, and m is a dummy variable equalling zero with probability $\pi_0 = 1-s$, and equalling one with probability $\pi_1 = s$. δ is a scale parameter, while the intercept of the regression (β_0) serves as a location parameter for the heterogeneity distribution.

In this model correlations between pairs of composite disturbances, $e_{it} = \alpha_{it} + u_{it}$ and $e_{it'} = \alpha_i + u_{it'}$, are constrained to be nonnegative. The variance of m is $Ms(1-s)$, and the variance of u is $\pi^2/3$; it can be shown that the correlation r between e_{it} and $e_{it'}$ equals

$$\frac{3\delta^2 Ms(1-s)}{3\delta^2 Ms(1-s) + \pi^2}$$

given the independence of m and u_{it} . r approaches zero as $s \rightarrow 0$ (or 1) and approaches one as δ (or M) $\rightarrow \infty$. The restriction to nonnegative correlations is not, however, troublesome in the context of repeated outcomes influenced by a time-invariant unmeasured attribute.

In panel data the observables consist of the binary outcomes d_{i1}, \dots, d_{iT} and the exogenous variables x_{i1}, \dots, x_{iT} . Let p_{itm} represent the conditional probability that $d_{it} = 1$ given $m = m$; that is

$$p_{itm} = \exp(\delta m + \beta' x_{it}) [1 + \exp(\delta m + \beta' x_{it})]^{-1} .$$

Then the conditional probability of the sequence d_{i1}, \dots, d_{iT} is

$$C_{im} = \prod_{t=1}^T p_{itm}^{d_{it}} q_{itm}^{1-d_{it}} ,$$

and, employing the parameters of the mixing distribution the unconditional probability of the sequence d_{i1}, \dots, d_{iT} is

$$U_i = \sum_{m=0}^M \pi_m \left(\prod_{t=1}^T p_{itm}^{d_{it}} q_{itm}^{1-d_{it}} \right) \quad (2)$$

or more compactly $U_i = \sum_{m=0}^M \pi_m C_{im}$.

In view of (2) the log-likelihood of the data is

$$\begin{aligned} \mathcal{L}_\theta &= \sum_{i=1}^N \ln \left[\sum_{m=0}^M \pi_m \left(\prod_{t=1}^T p_{itm}^{d_{it}} q_{itm}^{1-d_{it}} \right) \right] \\ &= \sum_{i=1}^N \ln \left(\sum_{m=0}^M \pi_m C_{im} \right) \end{aligned} \quad (3)$$

where θ is the parameter set $\{s, \delta, \beta_0, \beta_1\}$; for reasons which will immediately become clear we separate β into components β_0 (the intercept) and β_1 (the vector of coefficients on exogenous variables). Note that if for fixed $X = x_{11}, \dots, x_{it}, \dots, x_{NT}$ the function \mathcal{L}_θ is maximized by the set $\{\hat{s}, \hat{\delta}, \hat{\beta}_0, \hat{\beta}_1\}$ then it is also maximized by the set $\{1-\hat{s}, -\hat{\delta}, \hat{\beta}_0 + M\hat{\delta}, \hat{\beta}_1\}$, which produces an identical heterogeneity distribution. Thus without further loss of generality we impose the constraint $\delta \geq 0$.

The first-order conditions satisfied by the ML estimates of θ are as follows:

$$\frac{\partial \mathcal{L}}{\partial s} = \sum_{i=1}^N U_i^{-1} \left[\sum_{m=0}^M \pi'_m C_{im} \right] = 0; \quad (4a)$$

$$\frac{\partial \mathcal{L}}{\partial \delta} = \sum_{i=1}^N U_i^{-1} \left[\sum_{m=0}^M \pi_m C_{im} \left(\sum_{t=1}^T d_{it} q_{itm} - \sum_{t=1}^T (1-d_{it}) p_{itm} \right) \right] = 0; \quad (4b)$$

and

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \sum_{i=1}^N U_i^{-1} \left[\sum_{m=0}^M \pi_m C_{im} \left(\sum_{t=1}^T x_{ijt} d_{it} q_{itm} - \sum_{t=1}^T x_{ijt} (1-d_{it}) p_{itm} \right) \right] = 0 \quad (4c)$$

with $x_{i1t} \equiv 1$. In (4a), $\pi'_m = \frac{\partial}{\partial s} \pi_m = \pi_m \left[\frac{m - Ms}{s(1-s)} \right]$.

In the binary case ($N=1$) equation (4a) becomes particularly simple:

$$\frac{\partial \mathcal{L}}{\partial s} = \sum_{i=1}^N U_i^{-1} (C_{i1} - C_{i0}).$$

SAMPLING EXPERIMENTS

In this section are reported the results of several sampling experiments designed to explore the performance of the ML estimates of the binomial-mixture random-effects logit model. The experiments are not exhaustive, but illustrative. In each experiment 50 samples were drawn for fixed X and true parameter set θ^* , and with $M=1$ - heterogeneity in

the form of an unmeasured dummy variable. A single normally-distributed covariate was used in all experiments², and in all cases $\beta_0 = 0$ while $\beta_1 = 1$. Finally, in each experiment $NT = 1200$, so that the total number of observed outcomes is constant.

In implementing the model, transformations of the parameters s and δ were employed in order that restrictions on the arguments of the likelihood function could be relaxed. In particular, the likelihood function was maximized with respect to the transformed parameter set $\theta' = \left\{ \ln \left[\frac{1-s}{s} \right], \ln \delta, \beta_0, \beta_1 \right\}$.

Results of the experiments are presented in Table 1. From the most basic standpoint – the ability to obtain a solution – in almost all cases the model performed well, with the exceptions of experiments 7 and 10, representing extreme cases – of panel length ($T=2$) and the magnitude of omitted-variable effects ($\delta = 0.25$), respectively. For each of these experiments there were several samples for which the model failed to converge, and which therefore were discarded.³

Experiments 1-3 and 4-6 explore variations in s , which is the mean of the unmeasured dummy variable, for situations in which a moderate number (6) and a large number (12) of panel observations are available. Experiments 7 and 8 employ shorter panels (2 and 3 periods, respectively) while experiments 9 and 10 vary the magnitude of omitted-variable effects.

Table 1 reports two statistics for each parameter estimated: bias, which equals $\Sigma(\hat{\theta}_j - \theta_j^*)/50$, and root mean squared error (RMSE), which equals $[\Sigma(\hat{\theta}_j - \theta_j^*)/50]^{1/2}$. For the parameters s and δ , the average value, after inverting the transformation used in the actual estimation, is also shown. The bias of the estimators of all four parameters appears to be nonsystematic: in no case is the bias uniformly positive

² In particular, for each individual $x_{it} = \omega_{i0} + \omega_{it}$, where $\omega_0, \omega_1, \dots, \omega_T$ are *iid* standard normal variates. The resulting correlation between x_t and x_t^* is not an unrealistic situation in actual data.

³ It should be noted, however, that only a single set of starting values (chosen at random from the uniform distribution on the $[-1, 1]$ interval) was attempted for each sample; for the discarded samples it may have been possible to achieve convergence with a different set of starting values.

or negative. In applications the parameters of greatest concern are the regression coefficients β_0 and β_1 . For β_1 (represented by the parameter θ_4') the results are quite comforting, as the bias is small, nearly always less than one percent of the true value. In eight of the 10 experiments the bias of β_0 (represented by the parameter θ_3') exceeds that of β_1 by several times. The bias of $\theta_1' (= \ln \frac{l-s}{s})$ and $\theta_2' (= \ln \delta)$ is generally large, but, again, non-systematic.

In terms of RMSE, the results are similar to those just described: β_1 is estimated rather well, β_0 somewhat less well, and both s and δ are estimated considerably less well. Note also that in experiment 10, where δ equals only 0.25, the bias in the estimator of $\ln \delta$ is only about 10 percent of the true value, while the average of $\hat{\delta}$ is almost three times too high, a consequence of the nonlinearity of the transformation used, in combination with the sampling distribution of the estimator in this particular experiment.

A few tentative conclusions can be advanced on the basis of the results given in Table 1. First, the estimator of β_1 , which is perhaps of greatest importance in any application, is relatively unaffected by the variations explored here. Second, considering the whole parameter set, better estimates are obtained from long rather than short panels (cf. experiments 2, 5, 7 and 8), and when s is close to 0.5 (cf. experiments 1-3 or 4-6). And, finally, estimates of the parameters representing heterogeneity - s and δ - are relatively imprecise, particularly when δ takes on progressively more extreme values (cf. experiments 5, 9, and 10). None of these conclusions are very surprising, which is reassuring. Imprecision of ML estimators of the parameters of mixing distributions is frequently encountered, as for example in recent studies of heterogeneity in continuous-time econometric duration models (see, for example, Heckman and Singer, 1985).

CONCLUDING REMARKS

This note has proposed a random-effects approach to dealing with unmeasured variables in a panel logit model, using a binomial distribution as an assumption regarding the distribution of the unobservable. Monte-Carlo experiments suggest that maximum-likelihood estimation of the parameters of the model produces satisfactory results, at least for the structural parameters of greatest interest, and in the case where the data are generated by the assumed model. It remains to be seen how the ML approach would perform when the model is incorrectly specified, or, more importantly, when confronted by nonexperimental data.

We close with a few additional remarks on possible extensions or further variations on the basic model described:

- (1) It is not necessary that $T_i = T$ for all i ; each individual can be present for a different number of periods in the panel data, provided that the sampling plan is *noninformative* (see Hoem, 1985). That is, variation in T must not reveal information on the underlying process.
- (2) The model can be used to analyze discrete-time duration data (see Allison, 1982); unmeasured heterogeneity in a failure-time model can be a source of serious errors of inference, and a discrete-heterogeneity approach such as that outlined above can be used as one possible way to generalize a discrete-time duration model.
- (3) Distributions other than the binomial could, of course, be proposed as the mixing distribution. Distributions with infinitely many support points (such as the Poisson or negative binomial) would require additional study. Discrete mixing distributions have the attractive property of leading to mathematically convenient likelihood functions and gradients, as shown here; continuous mixing distributions might also be investigated.
- (4) Throughout, we have considered only binary outcomes, but the approach can easily be generalized to multinomial outcomes. In principle, if there are K outcome

categories, then we can identify the effects of $K-1$ unmeasured traits or "factors", each of which has a distinct "loading" - δ_{jk} ; $j, k = 1, \dots, K-1$ - in each of $K-1$ latent index functions.

- (5) Another attractive feature of the use of a discrete distribution to represent unmeasured heterogeneity is that it makes tractable the estimation of interaction effects between the unmeasured variable and measured coveriates; in other words, we could estimate models of the form

$$y_{it} = \delta m_i + \beta' x_{it} + \Gamma' m_i x_{it} + u_{it}$$

this suggestion, like those preceding it, is a subject for further study.

REFERENCES

- Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories". In *Sociological Methodology 1982*, ed. S. Leinhardt. San Francisco: Jossey-Bass Publishers.
- Heckman, James J. and Burton Singer. 1985. "Social Science Duration Analysis". In *Longitudinal Analysis of Labor Market Data*, ed. J.J. Heckman and B. Singer. Cambridge: Cambridge University Press.
- Hoem, Jan M. 1985. "Weighting, Misclassification, and Other Issues in the Analysis of Survey Samples of Life Histories". In *Longitudinal Analysis of Labor Market Data*, ed. J.J. Heckman and B. Singer. Cambridge: Cambridge University Press.
- Maddala, G. S. 1987. "Limited Dependent Variable Models Using Panel Data". *Journal of Human Resources* 22:305-338.
- Montgomery, Mark R., Toni Richards and Henry I. Braun. 1986. "Child Health, Breast-Feeding, and Survival in Malaysia: A Random-Effects Logit Approach" *Journal of the American Statistical Association* 81: 297-309.
- Rosenbaum, P. R. and D. B. Rubin. 1983. "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome". *Journal of the Royal Statistical Society (Series B)* 45: 212-218.
- Zeger, Scott L., Kung-Yee Liang and Steven G. Self. 1985. "The Analysis of Binary Longitudinal Data with Time-Independent Covariates". *Biometrika* 72: 31-38.

Table 1
Results from Sampling Experiments

	Experiment									
	1	2	3	4	5	6	7	8	9	10
T	6	6	6	12	12	12	2	3	12	12
N	200	200	200	100	100	100	600	400	100	100
s	0.1	0.3	0.5	0.1	0.3	0.5	0.3	0.3	0.3	0.3
δ	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	0.25
\hat{s}	.146	.300	.490	.113	.306	.506	.255	.262	.367	.398
$\hat{\delta}$	1.952	2.074	2.034	1.974	1.965	1.993	2.567	2.292	1.128	.749
True values ^a										
θ_1	-2.197	-.847	0	-2.197	-.847	0	-.847	-.847	-.847	-.847
θ_2	.693	.693	.693	.693	.693	.693	.693	.693	0	-1.386
Bias										
θ_1	.215	-.032	-.040	.081	.018	.022	-.325	-.256	.239	.025
θ_2	-.080	.029	.011	-.032	-.024	-.007	.197	.114	.091	.146
θ_3	-.034	-.003	.020	-.019	-.011	-.022	.042	.027	-.101	-.193
θ_4	.015	-.017	-.032	.002	.007	-.008	.018	.008	.029	.012
RMSE										
θ_1	.829	.395	.288	.398	.215	.169	.724	.611	1.061	2.137
θ_2	.346	.126	.104	.199	.117	.084	.371	.237	.264	3.913
θ_3	.143	.149	.140	.095	.098	.141	.203	.186	.269	.303
θ_4	.079	.075	.076	.079	.076	.089	.087	.076	.098	.058

^a In all experiments the true values for θ_3 and θ_4 are 0 and 1, respectively.