Political Analysis, 9:4

Bias in Conditional and Unconditional Fixed Effects Logit Estimation

Ethan Katz
Center for Basic Research in the Social Sciences,
Harvard University, 34 Kirkland Street,
Cambridge, Massachusetts 02138
e-mail: ekatz@post.harvard.edu

Fixed-effects logit models can be useful in panel data analysis, when N units have been observed for T time periods. There are two main estimators for such models: unconditional maximum likelihood and conditional maximum likelihood. Judged on asymptotic properties, the conditional estimator is superior. However, the unconditional estimator holds several practical advantages, and therefore I sought to determine whether its use could be justified on the basis of finite-sample properties. In a series of Monte Carlo experiments for $T < 20$, I found a negligible amount of bias in both estimators when $T \geq 16$, suggesting that a researcher can safely use either estimator under such conditions. When $T < 16$, the conditional estimator continued to have a very small amount of bias, but the unconditional estimator developed more bias as T decreased.

1 Introduction

SUPPOSE THAT WE observe N units for T time periods and that at each observation we record whether an event occurs. For example, the units might be country dyads, the time periods years, and the event war between the two countries in the dyad. Let $\mathbf{y}$ denote the $NT \times 1$ vector containing these binary observations. Suppose, further, that we believe that the probability of an event is a function of one or more covariates. Let $\mathbf{X}$ denote the $NT \times k$ matrix that consists of a column of ones concatenated with the $NT \times (k-1)$ matrix of observations on the $(k-1)$ covariates. One way to model the relationship between $\mathbf{y}$ and $\mathbf{X}$ is to use a fixed-effects logit specification, as follows:

$$P(y_{i,t} = 1 \mid \alpha_i, \boldsymbol{\beta}) = \frac{\exp(\alpha_i + \mathbf{x}_{i,t}\boldsymbol{\beta})}{1 + \exp(\alpha_i + \mathbf{x}_{i,t}\boldsymbol{\beta})}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T \quad (1)$$

This model makes the following assumptions: first, conditional on $\mathbf{x}_{i,t}$, $y_{i,t}$ is an independent Bernoulli random variable with probability given by (1); second, $P(y_{i,t} = 1)$ depends on $\mathbf{x}_{i,t}$

Author's note: I owe a tremendous debt to Gary King, who suggested this topic in the first place and who provided advice and encouragement at every stage. I thank two anonymous referees for helpful comments and Donald P. Green for sharing his data. These data, as well as programs to carry out the analyses described in this paper, are available at the *Political Analysis* web site.

Copyright 2001 by the Society for Political Methodology

through the logistic function; third, $P(y_{i,t} = 1)$ is governed by both a vector β of k common ("structural") parameters and a unit-specific ("incidental") parameter α_i. For a discussion of the application of such models to international relations, see Green et al. (2001), Beck and Katz (2001), and King (2001).

One way to estimate the parameters of this model is to include dummy variables for the units and then to maximize the unconditional likelihood function. The unconditional maximum-likelihood estimator of the incidental parameters is consistent as $T \to \infty$ for fixed N but inconsistent as $N \to \infty$ for fixed T. The inconsistency arises because the number of incidental parameters increases without bound, while the amount of information about each incidental parameter remains fixed (Neyman and Scott 1948). The unconditional estimator of the structural parameters is also, in general, inconsistent (for examples see Andersen 1973).

However, Chamberlain (1980; see also Andersen 1970) offers an estimator of the structural parameters that is consistent even in the presence of incidental parameters. This estimator is obtained by conditioning the likelihood function on minimal sufficient statistics for the incidental parameters and then maximizing the conditional likelihood function; in the logit case, such statistics can be $\sum_{t=1}^{T} y_{i,t}$.

Although asymptotic results favor the conditional estimator, an applied researcher might prefer unconditional maximum likelihood for several practical reasons. First, the conditional estimator is more computationally complex. Second, it cannot produce any estimates of the incidental parameters, which are necessary to calculate quantities of interest such as absolute risks, risk ratios, and risk differences (for elaboration of this point see King 2001). Third, missing data are especially problematic for the conditional estimator; since it conditions on the total number of events for a unit, the loss of even one observation within a unit means that the entire unit must be dropped. Fourth, independent variables that do not vary by time cancel out in the likelihood function, and thus their effects cannot be estimated.

Therefore, I seek to determine whether use of the unconditional estimator can be justified based on finite-sample properties. There has been little research into the finite-sample properties of either estimator. Wright and Douglas (1977) perform a Monte Carlo study of relative bias in the unconditional estimator, and they find bias ranging from 1% of the parameter value, when T is 80, to 6%, when T is 20. Heckman (1981) performs a Monte Carlo study of bias in the unconditional estimator of a fixed-effects probit model. He considers the case when T is 8 and finds a very small bias; when the true parameter is 1.0, the average estimate is 0.91, and when the true parameter is 0.1, the average estimate is 0.097.

I aim to fill the gaps in these studies by offering a more complete guide to bias in the estimation of fixed-effect logit models when $T < 20$. Such knowledge is generally useful in political methodology, and in particular, it will help researchers choose between the unconditional and the conditional estimators.

2 Methods

Bias is defined as $E(\hat{\beta}) - \beta$, where $\hat{\beta}$ denotes the estimate of the parameter β. Therefore, my goal was to calculate $E(\hat{\beta})$ for different values of T. I estimated $E(\hat{\beta})$ through simulations: For a fixed value of X, I drew repeated random samples from the conditional distribution of Y, estimated β for each sample, and averaged the estimates across the samples. The average of $\hat{\beta}$ is denoted $\hat{E}(\hat{\beta})$, since it approximates $E(\hat{\beta})$. In this paper, a simulation refers to a single draw of Y; an experiment refers to a set of M simulations across which $\hat{\beta}$ is

averaged. I studied the relationship between $\hat{E}(\hat{\beta})$ and T for different configurations of N, β, and the proportion of ones in the Y vector. I included the proportion of ones, denoted p, to determine the effect of rare events.

The specific data-generating process was as follows: I drew each $y_{i,t}$ from a Bernoulli distribution governed by

$$P(y_{i,t} = 1 \mid \gamma, \alpha_i, \beta) = \frac{\exp(\gamma + \alpha_i + x_{i,t}\beta)}{1 + \exp(\gamma + \alpha_i + x_{i,t}\beta)} \tag{2}$$

I controlled p by adjusting the constant term, γ, and I set α_i to 1 for all i. When $T > 2$, I drew the scalar $x_{i,t}$ (once, for the entire experiment) from a uniform distribution on (0, 1); when T was 2, $x_{i,1} = 0$ and $x_{i,2} = 1$ for all i. I drew the values of N, β, and p from the following sets:

$$\begin{aligned} N &\in \{50, 100\} \\ \beta &\in \{-1.0, -0.5, -0.1, 0.1, 0.5, 1.0\} \\ p &\in \{0.90, 0.56, 0.32, 0.15, 0.06\} \end{aligned}$$

For each configuration, I varied T from 2 to 16. Each experiment consisted of 500 simulations.

In each experiment, I calculated $\hat{E}(\hat{\beta})$ as follows:

$$\hat{E}(\hat{\beta}) = \frac{1}{M} \sum_{j=1}^{M} \hat{\beta}_j \tag{3}$$

I calculated the standard error of the estimate, the error from using only a finite number of samples to estimate the mean of the sampling distribution of $\hat{\beta}$, as follows:

$$\sqrt{Var(\hat{E}(\hat{\beta}))} = \sqrt{\frac{(1/(M-1)) \sum_{j=1}^{M} (\hat{\beta}_j - \hat{E}(\hat{\beta}))^2}{M}} \tag{4}$$

3 Results

When T is 2, X can sometimes predict Y perfectly, in which case β is not identified. Nonetheless, the general pattern is clear and systematic. The average conditional estimates are always very close to the true values. The average unconditional estimates are approximately 2β when T is 2 but quickly come close to the true values as T increases, so that when T reaches 16, the unconditional estimator does as well as the conditional estimator. The case when T is 2 accords with asymptotic results, since Andersen (1973, p. 66) proves that $\hat{\beta} \to 2\beta$ as $N \to \infty$.

Figure 1 displays a sample of the results graphically. It contains only one case ($N = 100$, $\beta = 0.5$, $p = 0.56$), since the pattern is essentially the same in all cases. Table 1 shows how $\hat{E}(\hat{\beta})$ varies with T for different configurations of N and β; p is fixed at 0.90. Only the results for $\beta = 0.1$ and 1.0 are shown; the other configurations are omitted because they display almost identical patterns.

Table 1 How $\hat{E}(\hat{\beta})$ varies with N, T, and β[a]

	$\beta = 0.1$		$\beta = 1.0$	
	$N = 100$	$N = 50$	$N = 100$	$N = 50$
$T = 2$				
Cond	0.097 (0.021)	*	*	*
Uncond	0.194 (0.043)	*	*	*
$T = 4$				
Cond	0.097 (0.028)	0.098 (0.043)	0.988 (0.035)	1.104 (0.051)
Uncond	0.130 (0.037)	0.130 (0.059)	1.341 (0.048)	1.523 (0.072)
$T = 8$				
Cond	0.126 (0.019)	0.089 (0.025)	0.986 (0.023)	1.042 (0.033)
Uncond	0.144 (0.022)	0.102 (0.028)	1.130 (0.028)	1.198 (0.038)
$T = 12$				
Cond	0.090 (0.015)	0.087 (0.021)	1.017 (0.018)	0.965 (0.025)
Uncond	0.098 (0.016)	0.095 (0.023)	1.111 (0.020)	1.056 (0.027)
$T = 16$				
Cond	0.096 (0.013)	0.131 (0.018)	1.002 (0.016)	0.965 (0.022)
Uncond	0.103 (0.014)	0.139 (0.019)	1.069 (0.017)	1.031 (0.024)

[a]Cond, conditional maximum-likelihood estimate; uncond, unconditional maximum-likelihood estimate. Asterisks indicate that X predicted Y perfectly. Monte Carlo standard errors are given in parentheses. p is approximately 0.90. The number of simulations is 500.

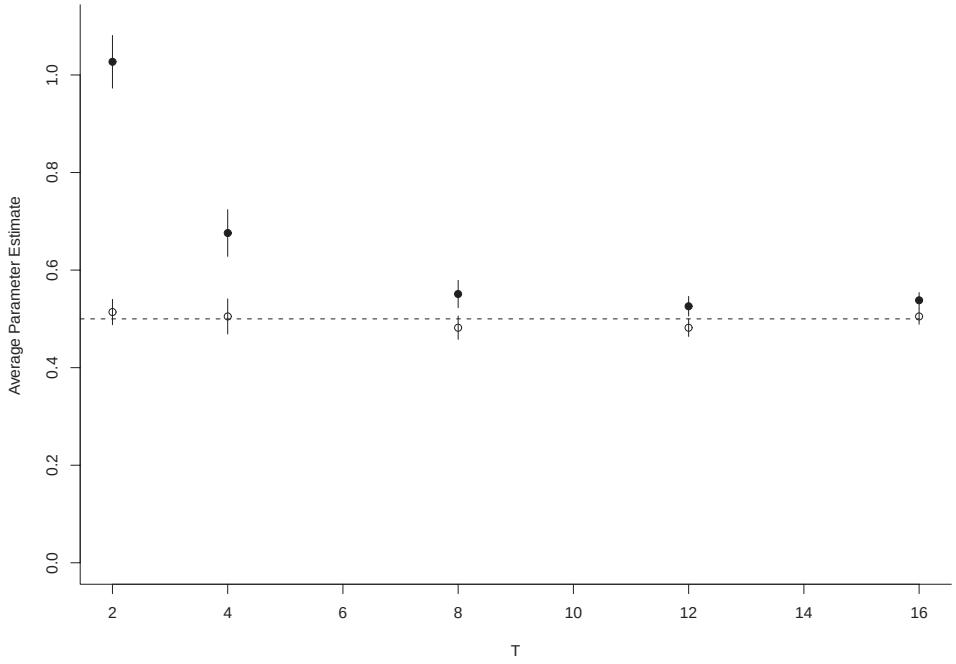

Fig. 1 How $\hat{E}(\hat{\beta})$ varies with T. Estimates of $E(\hat{\beta})$, along with 95% confidence intervals, are shown for $T = 2, 4, 8, 12$, and 16. Filled circles denote unconditional estimates and open circles denote conditional estimates. The dashed line marks the true value of β. N is set to 100, β to 0.5, and p to 0.56. The number of simulations is 500.

Table 2 How $\hat{E}(\hat{\beta})$ varies with T and p[a]

	$p = 0.56$ ($\gamma = -1$)	$p = 0.32$ ($\gamma = -2$)	$p = 0.15$ ($\gamma = -3$)	$p = 0.06$ ($\gamma = -4$)
$T = 2$				
Cond	0.514 (0.013)	0.506 (0.014)	0.486 (0.018)	*
Uncond	1.027 (0.027)	1.012 (0.029)	0.972 (0.037)	*
$T = 4$				
Cond	0.505 (0.018)	0.538 (0.020)	0.500 (0.028)	0.518 (0.042)
Uncond	0.676 (0.024)	0.722 (0.027)	0.670 (0.037)	0.699 (0.058)
$T = 8$				
Cond	0.482 (0.012)	0.510 (0.013)	0.515 (0.017)	0.489 (0.026)
Uncond	0.551 (0.014)	0.583 (0.015)	0.589 (0.012)	0.561 (0.029)
$T = 12$				
Cond	0.482 (0.009)	0.503 (0.010)	0.494 (0.013)	0.471 (0.019)
Uncond	0.526 (0.010)	0.548 (0.011)	0.539 (0.014)	0.515 (0.021)
$T = 16$				
Cond	0.505 (0.008)	0.525 (0.009)	0.491 (0.011)	0.496 (0.017)
Uncond	0.538 (0.008)	0.560 (0.010)	0.524 (0.012)	0.600 (0.018)

[a]Cond, conditional maximum-likelihood estimate; uncond, unconditional maximum-likelihood estimate. Asterisks indicate that X predicted Y perfectly. Monte Carlo standard errors are given in parentheses. N is 100 and β is 0.5. The number of simulations is 500.

Table 2 shows how $\hat{E}(\hat{\beta})$ varies with T for different values of p; N is fixed at 100 and β at 0.5. The pattern is essentially the same as in Table 1; the only change as p decreases is that the standard errors increase.

4 Example: Explaining War

In this section I compare the conditional and unconditional estimators using a data set on interstate conflict. Green et al. (2001) fit a fixed-effects logit model to a panel data set in which the units are country dyads, the time periods are years, and the dependent variable is whether war occurs in a dyad-year. The independent variables include measurements of democracy and bilateral trade.[1]

In the data set, N is 3075; T, which varies by dyad, ranges from 20 to 42 (each dyad is observed for some or all of the period 1951–1992); p is 0.01; and the total number of observations is 93,755.

I estimated coefficients using both the conditional and the unconditional estimators, and Table 3 displays the results. Since $T > 16$, the Monte Carlo findings suggest that the conditional and unconditional estimates should be very close, and this is indeed the case.

5 Discussion

The conditional estimator has a negligible amount of bias for all $T < 20$, and the unconditional estimator has a negligible amount of bias for $16 \leq T < 20$. The bias in the unconditional estimator grows as T decreases, reaching approximately 2β when T is 2.

[1]See Green et al. (2001) for a detailed description of the variables.

Table 3 Comparison of conditional and unconditional coefficient estimates for data on military conflict[a]

Covariate	*Conditional estimate*	*Unconditional estimate*
Alliance	−1.066 (0.426)	−1.115 (0.437)
Contiguity	1.902 (0.336)	1.993 (0.347)
Growth	−0.059 (0.012)	−0.061 (0.012)
Democracy	−0.003 (0.015)	−0.003 (0.015)
Capability ratio	0.387 (0.139)	0.401 (0.142)
Bilateral trade	−0.072 (0.186)	−0.074 (0.190)

[a]Standard errors are given in parentheses. N is 3075, T ranges from 20 to 42, and p is 0.01.

The implication for applied researchers is that the conditional estimator is always safe when $T < 20$ and that the unconditional estimator is safe when $16 \leq T < 20$. When $8 < T < 16$, the bias in the unconditional estimator is small and may be acceptable to the researcher, but when T is close to 2, the bias is substantial. Even in this case, however, the bias takes a predictable direction and magnitude.

References

Andersen, Erling Bernhard. 1970. "Asymptotic Properties of Conditional Maximum-Likelihood Estimators." *Journal of the Royal Statistical Society* 32:283–301.

Andersen, Erling Bernhard. 1973. *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygiejnisk Forsknings Institut.

Beck, Nathaniel, and Jonathan Katz. 2001. "Throwing Out the Baby with the Bath Water." *International Organization* 55:487–496.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47:225–238.

Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. "Dirty Pool." *International Organization* 55:441–468.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.

Hamerle, Alfred, and Gerd Ronning. 1995. "Panel Analysis for Qualitative Variables." In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, eds. Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York: Plenum Press.

Heckman, James. 1981. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process." In *Structural Analysis of Discrete Data*, eds. Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press.

King, Gary. 2001. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data." *International Organization* 55:497–507.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models.* New York: Chapman & Hall.

Neyman, J., and Elizabeth L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16:1–32.

Wright, Benjamin D., and Graham A. Douglas. 1977. "Conditional Versus Unconditional Procedures for Sample-Free Item Analysis." *Educational and Psychological Measurement* 37:573–586.