

On the Existence of Conditional Maximum Likelihood Estimates of the Binary Logit Model with Fixed Effects*

Martin Mugnier[†]

Abstract

By exploiting [McFadden \(1974\)](#)'s results on conditional logit estimation, we show that there exists a one-to-one mapping between existence and uniqueness of conditional maximum likelihood estimates of the binary logit model with fixed effects and the configuration of data points. Our results extend those in [Albert and Anderson \(1984\)](#) for the cross-sectional case and can be used to build a simple algorithm that detects spurious estimates in finite samples. As an illustration, we exhibit an artificial dataset for which the STATA's command `clogit` returns spurious estimates.

Keywords: separation, collinearity, binary choice models, fixed effects.

JEL Codes: C13, C25.

*I would like to thank Xavier D'Haultfœuille and Louis-Daniel Pape for their comments. All remaining errors are mine.

[†]CREST, martin.mugnier@ensae.fr.

1 Introduction

Suppose there are $i = 1, \dots, n$ individuals who are observed making a choice from $j \in \{0, 1\}$ alternatives, over $t = 1, \dots, T$ choice situations. Individual i 's sequence of choices is $Y_i := (Y_{i1}, \dots, Y_{iT})$ with elements $Y_{it} = \mathbb{1}\{i \text{ chooses } 1 \text{ in } t\}$ that indicate i 's choice in choice situation t . Suppose in addition to observe individual-specific covariates $X_{it} \in \mathbb{R}^p$ for individual i in choice situation t . Individual i 's sequence of covariates is then $X_i := (X'_{i1}, \dots, X'_{iT})'$. Let $\mathbf{Z}_n := (Y_i, X_i)_{1 \leq i \leq n}$ denote the sample data. Provided it exists and is unique, the conditional maximum likelihood estimator (that we denote by $\hat{\beta}_n^{\text{CMLE}}$) verifies

$$\hat{\beta}_n^{\text{CMLE}} := \arg \max_{\beta \in \mathbb{R}^p} \log L(\beta; \mathbf{Z}_n), \quad (1.1)$$

where

$$\log L(\beta; \mathbf{Z}_n) := \sum_{i=1}^N \log \frac{\exp\left(\sum_{t=1}^T Y_{it} X'_{it} \beta\right)}{\sum_{d_i: \sum_{t=1}^T (d_{it} - Y_{it}) = 0} \exp\left(\sum_{t=1}^T d_{it} X'_{it} \beta\right)}.$$

As [Andersen \(1970\)](#) showed (see also [Rasch, 1961](#), for an earlier reference), $\hat{\beta}_n^{\text{CMLE}}$ is consistent for the binary logit model with fixed effects. However, it may be the case that $\hat{\beta}_n^{\text{CMLE}}$ does not exist in finite samples (the maximum in (1.1) is not unique, or the conditional log-likelihood does not have a maximum). While [Albert and Anderson \(1984\)](#) established necessary and sufficient existence conditions for maximum likelihood estimates in cross-sectional logistic regression models,¹ such conditions for $\hat{\beta}_n^{\text{CMLE}}$ are lacking. Yet, these are of interest for at least two reasons. First, practitioners often use off-the-shelf programs with built-in nonlinear solvers which do not systematically detect, tag, or even report situations where estimates do *not* exist (see, e.g., [McCullough and Vinod, 2003](#)). Specifically, we show below with an artificial dataset that STATA's `clogit` stops after some iterations and returns spurious results, which further illustrates that the problem has not been widely reckognized.²

¹[Albert and Anderson \(1984\)](#) show that, under a rank condition on the design matrix, maximum likelihood estimates exist if and only if there is no binary linear classifier that perfectly predicts outcome from covariates for all data points lying outside the decision frontier.

²Noteworthy, the “`clogit`” section in the Stata User's Guide does not discuss the existence problem at all.

Second, it is easily seen that $\widehat{\beta}_n^{\text{CMLE}}$ does not take advantage of all the variation available in the data. This mechanically increases the probability of existence failure as we shall see. In this paper, we extend the data separation rules established in [Albert and Anderson \(1984\)](#) for cross-sectional logistic models. Our results hold under a rank condition which is similar to [Albert and Anderson \(1984, p. 2\)](#)' full rank assumption imposed on the matrix of covariates (see Assumption 1 below).

In Section 2, we derive our main existence theorem and provide a simple algorithm to detect existence failures. In Section 3, we exhibit an artificial dataset for which existence fails but `clogit` does not detect such failure and returns spurious estimates.

2 Necessary and Sufficient Existence Conditions

Following the terminology employed in [McFadden \(1974\)](#), the sample \mathbf{Z}_n results from a choice experiment composed of n distinct trials (X_i, B_i) , where B_i is the alternative set defined as

$$B_i := \left\{ d := (d_1, \dots, d_T) \in \{0, 1\}^T : \sum_{t=1}^T (d_t - Y_{it}) = 0 \right\}.$$

The alternative set contains $r_{ni} := \left(\sum_{t=1}^T Y_{it} \right)$ alternatives d_i^j , indexed by $j = 1, \dots, r_{ni}$, and with vectors of attributes $\sum_{t=1}^T d_{it}^j X_{it}$. Note that the number of alternatives, r_{ni} , differs from one individual to another. Let us define $r_n := \sum_{i=1}^N r_{ni}$. By rewriting Axiom 5 in [McFadden \(1974, p. 116\)](#) to fit our framework, we shall make the following assumption.

Assumption 1 *For all $\beta \in \mathbb{R}^p$, the $r_n \times p$ matrix whose rows are*

$$\left(\sum_{t=1}^T d_{it}^j X_{it} - \sum_{s=1}^{r_{ni}} \frac{\sum_{t=1}^T d_{it}^s X_{it} \exp(\sum_{t=1}^T d_{it}^s X_{it}' \beta)}{\sum_{\ell=1}^{r_{ni}} \exp(\sum_{t=1}^T d_{it}^\ell X_{it}' \beta)} \right)'$$

for $j = 1, \dots, r_{ni}$ and $i = 1, \dots, N$ is of rank p .

Assumption 1 holds when the data vary sufficiently across periods. A necessary condition is $r_n \geq p + N$. This is likely to hold in practice if p has a reasonable size. If $N \geq p$, it will hold generally since $r_{ni} \geq 2$, but it may also hold for $N < p$ if T is

large. The following condition is an adaptation of Axiom 6 in [McFadden \(1974, p. 116\)](#).

Assumption 2 *It does not exist $\beta^* \in \mathbb{R}^p \setminus \{0\}$ satisfying*

$$\sum_{t=1}^T (d_{it}^j - Y_{it}) X_{it}' \beta^* \geq 0$$

for $j = 1, \dots, r_{ni}$ and $i = 1, \dots, N$.

Assumption 2 is reminiscent of the separation and quasi-complete separation relationships (4) and (6) in [Albert and Anderson \(1984\)](#). However, it is specific to the panel data structure considered here.

Theorem 2.1 *Suppose Assumption 1 holds. Then, Assumption 2 is necessary and sufficient for the existence of a finite and unique $\hat{\beta}_n^{\text{CMLE}}$.*

Theorem 2.1 gives a sufficient and necessary condition for existence and uniqueness of conditional maximum likelihood estimates that depends only on the configuration of data points. It follows from an application of Lemma 3 in [McFadden \(1974\)](#). We now turn to the problem of finding an automated procedure for detecting if Assumption 2 holds in practice.

Theorem 2.2 *Suppose Assumption 1 holds. Then, Assumption 2 holds if and only if the minimum in the following quadratic programming problem is zero:*

$$\min_{u, \beta} u' u$$

subject to

$$u = \sum_{i=1}^N \sum_{j=1}^{r_{ni}} \beta_{ij} \sum_{t=1}^T (d_{it}^j - Y_{it}) X_{it} \text{ and } \beta_{ij} \geq 1.$$

Theorem 2.2 follows from an application of Lemma 4 in [McFadden \(1974\)](#). Note that a Python module called `BinLogitCMLE` that implements the program given in Theorem 2.2 before computing $\hat{\beta}_n^{\text{CMLE}}$ is made publicly available on the author's Github page.³

³<https://github.com/martinmugnier/BinLogitCMLE>.

3 An Example with Artificial Data

We generate an artificial dataset of 10 individuals (with personal identifiers \mathbf{id}) who are observed at periods $\mathbf{t} \in \{1, 2, 3\}$. The choice variable is $\mathbf{y}_{it} \in \{0, 1\}$ and there is a unique regressor $\mathbf{x}_{it} \in [0, 1]$. The data is displayed in Table 1.

Table 1: Artificial Data

\mathbf{id}	\mathbf{t}	\mathbf{x}_{it}	\mathbf{y}_{it}	\mathbf{id}	\mathbf{t}	\mathbf{x}_{it}	\mathbf{y}_{it}
1	1	0.48	0	6	1	0.33	0
1	2	0.50	0	6	2	0.62	1
1	3	0.42	0	6	3	0.48	0
2	1	0.52	1	7	1	0.001	0
2	2	0.46	0	7	2	0.87	1
2	3	0.60	1	7	3	0.85	1
3	1	0.57	1	8	1	0.78	1
3	2	0.58	1	8	2	0.95	1
3	3	0.39	0	8	3	0.51	1
4	1	0.40	0	9	1	0.26	0
4	2	0.37	0	9	2	0.99	1
4	3	0.52	1	9	3	0.43	0
5	1	0.10	0	10	1	0.17	0
5	2	0.41	0	10	2	0.22	0
5	3	0.25	0	10	3	1	1

Note that $\mathbf{y}_{it} = 1$ if and only if $\mathbf{x}_{it} > 0.5$. Hence, the stacked data is separated in [Albert and Anderson \(1984\)](#)' sense. Actually, it is easy to check that the data also violate Assumption 2. While the `logit` program detects the separation (see Figure 1), `clogit` does not detect violation of Assumption 2 and returns spurious estimates after three iterations (see Figure 2). Although the estimated standard error (resp. the log-likelihood function) is quite large (resp. almost zero), the user may miss the crucial fact that these results are informative only about the nonexistence of the estimate. We note that `BinLogitCMLE` detects the existence failure as expected.

References

- Albert, A. and Anderson, J. A. (1984), ‘On the existence of maximum likelihood estimates in logistic regression models’, *Biometrika* **71**(1), 1–10.
- Andersen, E. B. (1970), ‘Asymptotic properties of conditional maximum-likelihood estimators’, *Journal of the Royal Statistical Society. Series B (Methodological)* **32**(2), 283–301.
- McCullough, B. and Vinod, H. (2003), ‘Verifying the solution from a nonlinear solver: A case study’, *American Economic Review* **93**(3), 873–892.
- McFadden, D. (1974), Conditional logit analysis of qualitative choice behaviour, *in* P. Zarembka, ed., ‘Frontiers in Econometrics’, Academic Press New York, New York, NY, USA, pp. 105–142.
- Rasch, G. (1961), On general laws and the meaning of measurement in psychology, *in* ‘Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Medicine’, University of California Press, Berkeley, Calif., pp. 321–333.

4 Appendix

Figure 1: STATA's Logistic Regression

```
. logit y x  
  
outcome = x > .5 predicts data perfectly
```

Figure 2: STATA's Conditional Logit with Fixed Effects

```
. clogit y x, group(id)  
note: multiple positive outcomes within groups encountered.  
note: 3 groups (9 obs) dropped because of all positive or  
      all negative outcomes.  
  
Iteration 0:  log likelihood = -7.690286  
Iteration 1:  log likelihood =  2.132e-14  
Iteration 2:  log likelihood =  2.132e-14  
  
Conditional (fixed-effects) logistic regression  Number of obs   =      21  
                                                  LR chi2(1)           =     15.38  
                                                  Prob > chi2          =     0.0001  
Log likelihood =  2.132e-14                    Pseudo R2            =     1.0000
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	x	591.8446	8.55e+08	0.00	1.000	-1.68e+09 1.68e+09