Preprints.org

Article

Concavity of the Conditional Mle for Logit Panel Data Models with Imputed Covariates

Opeyo Peter Otieno [*] and Weihu Cheng

Article

Concavity of the Conditional Mle for Logit Panel Data Models with Imputed Covariates

Opeyo Peter Otieno [1,2,*] and Cheng Weihu [1]

[1]   Department of Statistics, Beijing University of Technology, Beijing 100124, China; chengweihu@bjut.edu.cn

[2]   Department of Statistics and Computational Mathematics, The Technical University of Kenya

*   Correspondence: opeyopeter@gmail.com

Abstract: In estimating logistic regression models, convergence of the maximization algorithm is critical, however, this may fail. Numerous bias correction methods for maximum likelihood estimates of the parameters have been conducted for cases of complete data sets and also longitudinal models. For binary response fixed effects panel data model, the conditional logit estimator is consistent for balanced data. When faced with missing covariates problem, researchers opt for various imputation techniques to make the data complete and without loss of generality consistent estimates still suffice asymptotically.   For maximum likelihood estimates of the parameters for logistic regression in cases of imputed covariates, the optimal choice of an imputation technique which yields the best estimates with minimum variance is still elusive. The main aim of this paper is to examine the behaviour of the Hessian matrix with optimal values of the imputed covariates vector which will make the Newton-Raphson algorithm to converge faster through a reduced absolute value of the product of the score function and the inverse fisher information component. We focus on a method used to modify the conditional likelihood function through partitioning of the covariate matrix. We also confirm from the moduli of the Hessian matrices that the log likelihood of a panel data logistic model has a global maximum as the parameter estimates.   Simulation results reveal that model based simulation perform better than classical imputation techniques yielding estimates with smaller bias and higher precision for the conditional maximum likelihood estimation of nonlinear panel models with single fixed effects.

Keywords: maximum likelihood; fixed effects; conditional logit; hessian matrix

1. Introduction

Parameter estimation is a key goal of inferential statistics and most researchers attempt to fit data into models that would produce the best of all possible parameter estimates. The motivation behind the parameter estimation is to make inferences about a study population using sample information and this calls for very well spelled out ways of ensuring that unbiased and precise estimates are achieved in every parameter estimation technique so applied. During sample data collection researchers encounter missing values in the study variables, a problem that leads to complications in statistical analyses through inaccurate estimates that may eventually lead to incorrect inferences and policy actions.

Specifically, when the response variable is binary, problems of missing covariates are further compounded by the nonlinear treatment of the model specification. Studies on missingness and parameter estimation have shown that frequentist techniques of imputation result into biased estimates with significant loss of power [1,2,3]. This problem cuts across every model with no exception to the logit model for binary choice response variables and several studies have made attempts to come up with reliable imputation techniques for missing observations so as to reduce the estimates' bias.

The logistic regression methods are often applied to procedures for describing the relationship between dichotomous outcome variables and other model covariates. The general method of estimating the logistic regression parameter is maximum likelihood and in a very general sense the maximum likelihood method yields values for the unknown parameters that maximize the probability of the observed set of data. Maximum likelihood (ML) method is occasionally prone to convergence problem, which occurs when the maximum likelihood estimates (MLE) do not exist.

This subject of assessing the behaviour of MLE for logistic regression model is important, as the applications of the logistic model stretch far and wide across research disciplines. There exist numerous works which discuss convergence problem on logistic regression model by Cox *et al.* [4] or the bias reduction by Firth D., Anderson J.A. and Richardson C. [5, 6]. Other studies outline many assumptions of the distributions ML estimators resulting from bias reduction technique, and impact of varying sample size to MLE [7, 8].

S. Lee [9] notes that statistical inference based on the logistic regression model is highly dependent on the asymptotic properties of the maximum likelihood estimator. This means that under large sample situations, the sampling distribution of the maximum likelihood (ML) estimators for the logistic regression coefficients is asymptotically unbiased and normal. Conversely, in small samples, the asymptotic properties of maximum likelihood estimations may not hold due to biased estimates [9, 10]. Firth's method has been introduced as one of the penalization techniques to correct or reduce the small sample bias of the ML estimators of the linear regression model [10, 11]. A comparison of the performance of the linear regression model based on maximum likelihood (ML) and Firth's penalized maximum likelihood estimation methods was done by S. Lee. The results showed that, as compared with penalized maximum likelihood, the LR model based on asymptotic ML worked slightly better in terms of statistical power although the difference in performance was not practically important [9]. The present paper centers to evaluate the susceptibility of the Hessian matrix to different imputation techniques by comparing the magnitudes of the determinants obtained from the Hessian of the log-likelihood function with the imputed covariate vector.

This section gives the background of the study by introducing the concept of panel data econometrics and mentions various approaches to panel data models' estimation where the conditional maximum likelihood estimation is mentioned as a solution to the incidental parameter problem in logistic panel data model.

Following the background introduction above, the rest of the paper is organized as follows. Section 2 focuses on the specification of the nonlinear binary choice panel data models. Section 3 highlights the incidental parameter problem in estimating logistic panel data model and shows how the conditional maximum likelihood approach circumvents it.   In section 4 we discuss the parameter estimation for a logit panel data model in which the covariate vector is partitioned into sample present values and missing or imputed values. This discerns the impact of missingness on the Hessian of the conditional maximum likelihood estimator of nonlinear binary choice logistic model. Monte Carlo simulation results are also given and discussed in this section to assess the impact of missingness on the determinant of the Hessian matrix and the parameter estimates. Finally, section 5 gives the summary and conclusions from the study and consequently provide recommendations for further study, based on the key findings from this work.

2. Model Specification

2.1. Panel Data

Observing experimental subjects or units over a repeated number of times collates a set of data referred to as panel data which provides two kinds of information – cross-sectional and time series. This unique characteristic allows panel data to account for individual differences that are time-invariant by applying regression techniques which allow us to adequately take advantage of the different types of information.

The logit panel data model then develops from the logistic regression to model binary choice response variables which have had wide applications in almost all research fields that conduct pretest and posttest studies with the aim of discerning the impact of the test. As such, for N units each observed T times, we have a total of $N \times T$ observations made.

2.2. The Logit Panel Data Model

Consider a population unit i is observed a time t for a binary response variable y against K explanatory variables to which we specify a general fixed effects panel data model as a special case of the generalized longitudinal model in the form

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \cdots + \beta_K x_{Kit} + c_i + u_{it}$$

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i + u_{it} \tag{1}$$

In model (1), $\boldsymbol{\beta}$ is a $1 \times k$ vector of parameters of the $k \times 1$ column vector $\boldsymbol{x}_{it}$. The parameter c_i captures all individual-specific time-invariant characteristics of the study population unit. For fixed effects (FE) models the parameter c_i is assumed to be correlated with $\boldsymbol{x}_{it}$ leaving only $u_{it} \sim N(0, \sigma_y^2)$ as the stochastic part of the model. Compounding c_i and u_{it} as a stochastic entity gives a random effects (RE) model. Let $y_{it} \in \{0,1\}$ for all i and t where the values 0 and 1 consecutively represent the failure and success of an event occurring as described in the research experiment. We then have y_{it} as a binary choice variable following a binomial distribution with probability of success for individual i at time t as $p_{it} = \Pr(y_{it} = 1) = E(y_{it}|\boldsymbol{x}_{it}, c_i) = F(\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i)$ and $E(y_{it}|\boldsymbol{x}_{it}, c_i) = 1 \times p_{it} + 0 \times (1 - p_i) = p_{it}$. This is so since $u_{it} \sim N(0, \sigma_y^2)$ by assumption, then strict exogeneity holds. The link function $F(\cdot)$ relates the binary outcome to the functional forms of the explanatory variables.

Under random sampling, $Pr(= 1|\boldsymbol{x}_{it}, c_i) = E(y_{it} = 1|\boldsymbol{x}_{it}, c_i; \beta)$ and if the binary response model above is correctly specified, we have the linear probability model (LPM) specified as

$$\left.\begin{array}{l} Pr(y_{it} = 1|\boldsymbol{x}_{it}, c_i) = \boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i \\[2mm] Pr(y_{it} = 0|\boldsymbol{x}_{it}, c_i) = 1 - (\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i) \end{array}\right\} \tag{2}$$

Adopting a linear probability model poses an absurdity of predicting "probabilities" of the response variable as either less than zero or greater than one. This shortfall is however addressed by specifying a monotonically increasing function F such that $F(.): \mathbb{R} \to [0,1]$ and

$$\left.\begin{array}{l} Pr(y_{it} = 1|\boldsymbol{x}_{it}, c_i) \to 1 \; as \; \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i \to \infty \\[2mm] Pr(y_{it} = 1|\boldsymbol{x}_{it}, c_i) \to 0 \; as \; \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i \to -\infty \end{array}\right\}$$

(3). This study adopts the logistic distribution as a nonlinear functional form of F as

$$F(\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i) = \frac{e^{\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i}}{1 + e^{\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i}} \tag{4}$$

which is between zero and one for all values of $\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i$. This is the cumulative distribution function (CDF) for a logistic variable whose parameters can be estimated. Equation (4) now represents what is known as the (cumulative) logistic distribution function which mitigates the limitations of LPM and has a total of $K + N$ parameters to be estimated as $\widehat{\boldsymbol{\beta}}$ and $\hat{c}_i$. It is verifiable that as $\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i$ ranges from $-\infty$ to $+\infty$, F ranges between 0 and 1 and that F is nonlinearly related to $\boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i$, thus satisfying the requirements for a probability function. F is nonlinear not only in X but also in the parameter vector $\boldsymbol{\beta}$ as can be seen clearly from (4) which implies that we cannot use the familiar OLS procedure to estimate the parameters. However, through linearization, this problem becomes more apparent than real. This is done by obtaining the odds ratio $\frac{Pr(y_{it} = 1|\boldsymbol{x}_{it}, c_i)}{Pr(y_{it} = 0|\boldsymbol{x}_{it}, c_i)}$ in favor of success i.e. the ratio of the probability that $y_{it} = 1$ to the probability that $y_{it} = 0$. It is realized that the logarithm of the odds ratio, is not only linear in X, but also (from the estimation viewpoint) linear in the parameters.

$$ln\left(\frac{Pr(y_{it} = 1|\boldsymbol{x}_{it}, c_i)}{Pr(y_{it} = 0|\boldsymbol{x}_{it}, c_i)}\right) = \boldsymbol{x}_{it}'\boldsymbol{\beta} + c_i \tag{5}$$

This log of the odds ratio (5) is called the logit, and hence the name logit model.

Among the approaches explored to estimate fixed effects models include: (a) Demeaning variables (b) Unconditional maximum likelihood or least squares dummy variables (LSDV) and (c) Conditional maximum likelihood estimation which is the most preferred method for logistic regressions.

The approaches used so far in estimating panel data models with fixed effects and continuous dependent variable y_{it} aim at controlling for these effects by eliminating their presence from the model and estimating the coefficients of the regressors. For categorical dependent variables, where specific nonlinear functions that preserve the structure of the dependent variable are considered, the conditional maximum likelihood estimation partials or 'conditions' the fixed effects out of the likelihood function by conditioning the probability of the regressand on the total number of events observed for each subject [12].

When the panel data sets are unbalanced due to cases of missing covariates, the estimation methods get computationally complicated and produce inefficient parameter estimates [13]. Various causes of missingness have been mentioned in literature among them being delayed entry, early exit or intermittent non-response from a study unit. As such, approaches suggested in literature on how to handle missing observations become valid in such cases. In this study we inspect the impact of missing data on the conditional maximum likelihood estimation procedures in nonlinear panel data models with an attempt to establish the best.

3. Incidental Parameter Problem and MLE

3.1. Incidental Parameter Problem

As specified in model (1), the presence of individual effects c_i complicates the computation of parameter estimates greatly, hence to obtain consistent estimates for the parameters for static linear models, we simply difference out the fixed effects. The number of parameters c_i increase with increase in sample size, a notion referred to as the incidental parameter problem realized by Neyman and Scott [14] and later reviewed by Lancaster [15]. For example, in the model (1) c_i and $\boldsymbol{\beta}$ are unknown parameters to be estimated and as $N \to \infty$ for fixed T, the number of parameters c_i increases with N . As such for nonlinear panel data models c_i cannot be consistently estimated for fixed T [16].

For linear panel data regression model with fixed T , only β can be estimated consistently by first getting rid of c_i using the within transformation. This is possible for the linear case because the MLE of β and c_i are asymptotically independent [17]. For qualitative binary choice model with fixed T, this is not possible as demonstrated by Chamberlain [12].

Hsiao [17] simply illustrates how the inconsistency of the ML estimate of c_i is transmitted into inconsistency for $\widehat{\boldsymbol{\beta}}_{mle}$. This is done in the context of a logistic model with one regressor x_{it} that is observed over two periods, with $x_{i1} = 0$ and $x_{i2} = 1$ where as $N \to \infty$ with $T = 2$, $plim\hat{\beta}_{mle} = 2\beta$. Greene [18] shows that despite the large number of incidental parameters, one can still force maximum likelihood estimation for the fixed effects model by including a large number of dummy variables.

3.2. The Unconditional Likelihood Function

The logistic model is estimated by means of Maximum Likelihood (ML), a technique that seeks a particular vector $\hat{\beta}^{ML}$ that gives the greatest likelihood of observing the outcomes in the sample $\{y_1, y_2, \dots\}$ conditional on the explanatory variables x.

By assumption, $p_{it} = \Pr(y_{it} = 1) = E(y_{it}|\boldsymbol{x_{it}}, c_i) = F(\boldsymbol{x'_{it}\beta} + c_i)$ and $1 - p_{it} = \Pr(y_{it} = 0) = 1 - F(\boldsymbol{x'_{it}\beta} + c_i)$ It then follows that the probability of observing the entire sample is

$$L(y|x;\beta) = \prod_{i=1}^{N}(F(\boldsymbol{x_{it}\beta} + c_i))^{y_i} [1 - F(\boldsymbol{x_{it}\beta} + c_i)]^{1-y_i} \tag{6}$$

The log likelihood function for the sample is

$$lnL(y|\boldsymbol{x};\beta) = \sum_{i=1}^{N}\{y_i lnF(\boldsymbol{x_{it}\beta} + c_i) + (1 - y_i)ln[1 - F(\boldsymbol{x_{it}\beta} + c_i)]\} \tag{7}$$

The MLE $\hat{\beta}^{ML}$ maximizes this log likelihood function (7).

3.3. Conditional Likelihood Function for Logistic Panel Data Model

If F is the logistic CDF then (7) gives the logistic log likelihood as:

$$lnL(y|\boldsymbol{x};\beta) = \sum_{i=1}^{N}\left\{y_i ln\left(\frac{e^{x_{it}\beta+c_i}}{1 + e^{x_{it}\beta+c_i}}\right) + (1 - y_i)ln\left(\frac{1}{1 + e^{x_{it}\beta+c_i}}\right)\right\} \tag{8}$$

The logistic model is preferred over the alternative probit model because it yields a consistent estimator of $\boldsymbol{\beta}$ without making any assumptions about how c_i is related to $\boldsymbol{x_{it}}$ except strict exogeneity. This is possible, because the logistic functional form enables us to eliminate c_i from the estimating equation, once we condition on the "minimal sufficient statistic" for c_i. As such we obtain the conditional likelihood function whose parameters are estimated.

Considering the conditional probabilities when $T = 2$, we have that:

$$Pr(y_{i1} + y_{i2} = 1|x_{i1}, x_{i2}, c_i,) = \frac{e^{x_{i1}\beta+c_i} + e^{x_{i2}\beta+c_i}}{[1 + e^{x_{i1}\beta+c_i}][1 + e^{x_{i2}\beta+c_i}]} \tag{9}$$

$$Pr(y_{i1} = 0, y_{i2} = 1|x_{i1}, x_{i2}, c_i, y_{i1} + y_{i2} = 1) = \frac{e^{x_{i2}\beta+c_i}}{e^{x_{i1}\beta+c_i} + e^{x_{i2}\beta+c_i}}$$

$$= \frac{e^{(x_{i2}-x_{i1})\beta}}{1 + e^{(x_{i2}-x_{i1})\beta}} \tag{10}$$

and

$$Pr(y_{i1} = 1, y_{i2} = 0|x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1) = \frac{1}{1 + e^{(x_{i2}-x_{i1})\beta}} \tag{11}$$

Note that conditioning is on $y_{i1} + y_{i2} = 1$, for which y_{it} changes between the two time periods ensures that the $c_i's$ are eliminated and therefore $\sum_t y_{it}$ is a sufficient statistic for the fixed effects.

Probabilities (10) and (11) are conditional on $y_{i1} + y_{i2} = 1$ and are independent of c_i.

The probability distribution function is expressed as

$$Pr(y_{i1}, y_{i2}|x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1) = \begin{cases} 1 & if\ (y_{i1}, y_{i2}) = (0,0)or(1,1) \\ \frac{1}{1+e^{(x_{i2}-x_{i1})\beta}} & if\ (y_{i1}, y_{i2}) = (1,0) \\ \frac{e^{(x_{i2}-x_{i1})\beta}}{1+e^{(x_{i2}-x_{i1})\beta}} & if\ (y_{i1}, y_{i2}) = (0,1) \end{cases} \tag{12}$$

The conditional log likelihood function from (8) is then given as

$$lnL = \sum_{i=1}^{N}\left\{d_{01i}ln\left(\frac{e^{(x_{i2}-x_{i1})\beta}}{1 + e^{(x_{i2}-x_{i1})\beta}}\right) + d_{10i}ln\left(\frac{1}{1 + e^{(x_{i2}-x_{i1})\beta}}\right)\right\} \tag{13}$$

where d_{01i} selects the individuals for which the dependent variable changed from 0 to 1 while d_{10i} selects the cases for which the dependent variable changed from 1 to 0.

Hence, by maximizing the conditional log likelihood function (13) we obtain consistent estimates of β, regardless of whether c_i and x_{it} are correlated. Generally, from an assessment of bias and root mean square errors, the conditional logit estimator performs better than the unconditional logit estimator when various imputation techniques are performed more so when the sample size is large [19].

The trick is thus to condition the likelihood on the outcome series (y_{i1}, y_{i2}), and in the more general case. For example, if T = 3, we can condition on $\sum_t y_{it} = 1$, with possible sequences (1,0,0) , (0,1,0) , (0,0,1) , or on $\sum_t y_{it} = 2$ with possible sequences (1,1,0) , (0,1,1) , (1,0,1). The general conditional probability of the response variable $(y_{i1}, y_{i2}, \ldots., y_{iT})$ given $\sum_t y_{it}$ is

$$Pr(y_{i1}, y_{i2}, \ldots, y_{iT} | \boldsymbol{X_i}, \textstyle\sum_t y_{it}) = \frac{e^{(\sum_t y_{it} x_{it} \boldsymbol{\beta})}}{\sum_{d \in B_i} e^{(\sum_t d_{it} x_{it} \boldsymbol{\beta})}} \tag{14}$$

where $B_i = \{(d_{i1}, d_{i2}, \ldots, d_{iT}) | d_{it} = 0,1 \ and \ \sum_t d_{it} = \sum_t y_{it}\}$

4. Parameter Estimation with Imputed Covariate Sub-matrix

4.1. Partitioned Covariate Matrix

In the presence of missing observations in the covariate vector $\boldsymbol{x_{it}}$, we express it as a sum of two vectors $\boldsymbol{x_{it_s}}$ and $\boldsymbol{x_{it_I}}$ for the sample-present covariate values and the missing covariate values respectively. Therefore, we have the conditional probabilities (10) and (11) as

$$Pr(y_{i1} = 0, y_{i2} = 1 | \boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, y_{i1} + y_{i2} = 1) = \frac{e^{\Delta x_{i_I} \boldsymbol{\beta}}}{e^{-\Delta x_{i_s} \boldsymbol{\beta}} + e^{\Delta x_{i_I} \boldsymbol{\beta}}} \tag{15}$$

and

$$Pr(y_{i1} = 1, y_{i2} = 0 | \boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, y_{i1} + y_{i2} = 1) = \frac{e^{-\Delta x_{i_s} \boldsymbol{\beta}}}{e^{-\Delta x_{i_s} \boldsymbol{\beta}} + e^{\Delta x_{i_I} \boldsymbol{\beta}}} \tag{16}$$

respectively, where $\Delta \boldsymbol{x_{i_I}} = (x_{i2_I} - x_{i1_I})$ and $\Delta \boldsymbol{x_{i_s}} = (x_{i2_s} - x_{i1_s})$.

The conditional log likelihood function is then obtained using equations (15) and (16) as

$$lnL = \sum_{i=1}^{N} \left\{ d_{01i} ln\left(\frac{e^{\Delta x_{i_I} \boldsymbol{\beta}}}{e^{-\Delta x_{i_s} \boldsymbol{\beta}} + e^{\Delta x_{i_I} \boldsymbol{\beta}}} \right) + d_{10i} ln\left(\frac{e^{-\Delta x_{i_s} \boldsymbol{\beta}}}{e^{-\Delta x_{i_s} \boldsymbol{\beta}} + e^{\Delta x_{i_I} \boldsymbol{\beta}}} \right) \right\} \tag{17}$$

where d_{01i} selects the individuals for which the dependent variable changed from 0 to 1 while d_{10i} selects the cases for which the dependent variable changed from 1 to 0.

Consistent estimates of the parameters of equation (17) are solved for by iterative technique using Newton-Raphson algorithm.

4.2. Newton-Raphson Algorithm and the Hessian Matrix Optimization of the Log likelihood function

Newton-Raphson method is an iterative procedure to calculate the roots of function f. In this method, we want to approximate the roots of the function by calculating

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{18}$$

where x_{n+1} are the $(n + 1)^{th}$ iteration. The goal of this method is to make the approximated result as close as possible with the exact result (that is, the roots of the function). If f is defined as the gradient function (score vector) then the first derivative of f gives the Hessian matrix which is the matrix of the second order derivatives of the likelihood function.

Starting from an initial estimate, $\boldsymbol{\beta}^{(0)}$, the algorithm consists of iterating the estimate at step h as

$$\boldsymbol{\beta}^{(h)} = \boldsymbol{\beta}^{(h-1)} + J(\boldsymbol{\beta}^{(h-1)})^{-1} s(\boldsymbol{\beta}^{(h-1)}) \tag{19}$$

where, $s(\boldsymbol{\beta}) = \frac{\partial \ln L}{\partial \boldsymbol{\beta}}$ is the score vector and $J(\boldsymbol{\beta}) = -\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is the observed information matrix obtained by computing and negating the Hessian matrix.

The score vector and observed Hessian matrix from the log likelihood function are respectively,

$$s(\boldsymbol{\beta}) = \frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left\{ d_{01i} \left[\Delta \boldsymbol{x'}_{i_I} - \left(\frac{-\Delta \boldsymbol{x'}_{i_s} e^{-\Delta x_{i_s} \boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_I} e^{\Delta x_{i_I} \boldsymbol{\beta}}}{e^{-\Delta x_{i_s} \boldsymbol{\beta}} + e^{\Delta x_{i_I} \boldsymbol{\beta}}} \right) \right] \right.$$

$$\left. - d_{10i} \left[\Delta \boldsymbol{x'}_{i_s} + \left(\frac{-\Delta \boldsymbol{x'}_{i_s} e^{-\Delta x_{i_s} \boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_I} e^{\Delta x_{i_I} \boldsymbol{\beta}}}{e^{-\Delta x_{i_s} \boldsymbol{\beta}} + e^{\Delta x_{i_I} \boldsymbol{\beta}}} \right) \right] \right\} \tag{20}$$

$$J(\boldsymbol{\beta}) = -\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{N} \left\{ d_{01i} \left[\left(\frac{\Delta x'_{i_s} \Delta x_{i_s} e^{-\Delta x_{i_s}\beta} + \Delta x'_{i_I} \Delta x_{i_I} e^{\Delta x_{i_I}\beta}}{e^{-\Delta x_{i_s}\beta} + e^{\Delta x_{i_I}\beta}} \right) - \boldsymbol{D}'\boldsymbol{D} \right] + \right.$$

$$\left. d_{10i} \left[\left(\frac{\Delta x'_{i_s} \Delta x_{i_s} e^{-\Delta x_{i_s}\beta} + \Delta x'_{i_I} \Delta x_{i_I} e^{\Delta x_{i_I}\beta}}{e^{-\Delta x_{i_s}\beta} + e^{\Delta x_{i_I}\beta}} \right) - \boldsymbol{D}'\boldsymbol{D} \right] \right\} \qquad (21)$$

where $\boldsymbol{D} = \left(\frac{-\Delta x'_{i_s} e^{-\Delta x_{i_s}\beta} + \Delta x'_{i_I} e^{\Delta x_{i_I}\beta}}{e^{-\Delta x_{i_s}\beta} + e^{\Delta x_{i_I}\beta}} \right)$

For well-defined parameter estimates of the log likelihood function, it is sufficient that (a) the log likelihood function must be concave indicating that the model is identified; (b) the Hessian matrix must be negative semi definite yielding a negative curvature of the log likelihood plot. This means that the determinant of the Hessian matrix, when evaluated at a critical point of a function, is equal to the Gaussian curvature of the function considered as a manifold. Concavity of the log-likelihood function is easily established when all eigenvalues of its Hessian are negative. Therefore, the necessary condition for a function to be concave is that the determinant of the Hessian matrix of the function should be greater than zero.

In this study we confirm that the conditional log likelihood function of the logit panel data model preserves its concavity even when different imputation techniques are applied to the missing covariates matrix X. Establishing the concavity or convexity of the log-likelihood function becomes a necessary condition to help know whether the solutions or parameter estimates are optimally local or global. For the nonlinear logit panel data model, the maximum likelihood estimates are yielded when the Hessian matrix is negative semi-definite resulting from a strictly concave log-likelihood function.

We use simulations to assess the relationship between the Hessian modulus and the properties of the parameter estimates for the conditional MLE of logit panel data model with various imputation techniques for missing covariates.

4.3. Monte Carlo Simulation

In this section, we present the results of Monte Carlo simulation to investigate the concavity of the log likelihood function through the behaviour of the hessian matrix when different imputation techniques are used to fill up for the missing covariates. Focusing on the conditional ML estimator for the logistic model given by the maximization of (17) the simulation results will compare the properties of the Hessian matrices of the conditional log likelihood function resulting from the new data sets obtained after imputation.

The simulation compares different sets of panel data generated by imputing covariates with imposed missingness patterns. This is achieved by imputing the missing observations and substituting the imputed vector $\boldsymbol{x}_{it_I}$ into the conditional log likelihood function (17) where the item based and model based imputation methods are used to fill up for the missing covariates. We consider a binary response variable that is specified by the relation model:

$$y_{it} = 1(\boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + v_{it} \geq 0) \quad i = 1,2,\dots,n \quad t = 1,2,\dots,T$$

$$y_{it} = 1\big(c_i + \beta_1 x^{(1)}{}_{it} + \beta_2 x^{(2)}{}_{it} + \beta_3 x^{(3)}{}_{it} + \beta_4 x^{(4)}{}_{it} + \beta_5 x^{(5)}{}_{it} + v_{it} \geq 0\big) \quad i$$

$$= 1,2,\dots,n \quad t = 1,2,\dots,T \qquad (22)$$

where $\boldsymbol{x}_{it}$ is a vector of five explanatory variables drawn from uniform, binomial and normal distributions (see Table 4.1) and the error term v_{it} has a logistic distribution given by $v_{it} = ln\left|\frac{u_{it}}{1+u_{it}}\right|$ with $u_{it} \sim N(0,1)$. β_1 to β_5 were fixed as β_1=1, β_2=-1, β_3=1, β_4=1 and β_5=1. The fixed effects c_i are obtained as functions of $x^{(1)}$ and T by the relation $c_i = \frac{\sqrt{T}\sum x^{(1)}}{n} + a_i$ with $a_i \sim N(0,1)$.

Table 4.1. Description of variables.

Variable	x_1	x_2	x_3	x_4	x_5
Type	continuous	continuous	continuous	discrete	discrete
Distribution	N~(0, 1)	U~ (0, 1)	N~ (0.5, 0.5)	B~(nT, 2, 0.65)	Binary

To establish the sample sizes we imposed an expected probability of success as $Pr(y_{it} = 1 | \boldsymbol{x_{it}}, c_i) = 0.5$ and plausible coefficients of variation (CoV) as 0.2, 0.14 and 0.09 respectively in the relation $N \cong \frac{1}{pr(y) \times (CoV)^2}$. In order to factor in small, medium and large sample size possibilities, three different values of N (N = 50, N =100 and N = 250) were used for all sets of data fitted into the models to enable in-depth comparisons and also assess the impact of sample size on the determinant of the Hessian matrix of the log-likelihood function. To evaluate the impact of the proportion of missingness we use two proportions of 10% and 30% by randomly deleting the desired proportion of observations from the data set and imputing them back for each sample size.

For each data set specified, we find the determinants of the Hessian matrices and plot against the corresponding data code for ease of comparison across sample sizes. We use the determinant of the Hessian matrix as a generalization of the second derivative test for single variable functions. As such if the determinant of the Hessian is positive we have an optimum or extreme value – a maximum if the matrix is negative (semi) – definite. This shows that the log likelihood is a concave function. The imputation techniques used herein are: mean imputation; median imputation; last value carried forward; multiple imputation with chained equations and Bayesian imputation.

Table 2. Parameter Estimates by Conditional MLE for Complete Unimputed Panel Data Set.

		Complete Unimputed Data				
		β1	β2	β3	β4	β5
10% Missingness	n=50	1.097358	-1.1006	1.162922	1.088189	1.135867
	n=100	1.063353	-1.04996	1.060443	1.042268	1.07011
	n=250	1.026624	-1.02073	1.020552	1.020266	1.006623
30% Missingness	n=50	1.118411	-1.06917	1.112667	1.122587	1.10587
	n=100	1.05757	-1.02386	1.074032	1.033003	1.052396
	n=250	1.027265	-1.01406	1.008524	1.015673	0.993208

Table 3. Parameter Estimates by Conditional MLE for Mean Imputed Panel Data Set.

		Mean Imputed Data				
		β1	β2	β3	β4	β5
10% Missingness	n=50	1.069795	-1.07215	1.12953	1.059192	1.245338
	n=100	1.032826	-1.00675	1.023471	1.006381	1.154124
	n=250	1.000421	-0.9956	0.988925	0.994209	1.120467
30% Missingness	n=50	1.043838	-0.966003	0.968262	1.008104	1.329941
	n=100	1.004478	-0.93376	0.970724	0.929846	1.262566
	n=250	0.980142	-0.93882	0.922916	0.913546	1.238532

Table 4. Parameter Estimates by Conditional MLE for LVCF Imputed Panel Data Set.

		LVCF Imputed Data				
		β1	β2	β3	β4	β5
10% Missingness	n=50	0.943744	-0.92664	0.977888	0.923632	1.126666
	n=100	0.895114	-0.87648	0.889337	0.878934	1.083107
	n=250	0.869159	-0.8563	0.853653	0.864803	1.03423
30% Missingness	n=50	0.682111	-0.60787	0.626598	0.665022	0.967133
	n=100	0.660872	-0.60855	0.620407	0.604995	0.92563
	n=250	0.633169	-0.59185	0.594952	0.592881	0.911108

Table 5. Parameter Estimates by Conditional MLE for Median Imputed Panel Data Set.

		Median Imputed Data				
		β1	β2	β3	β4	β5
10% Missingness	n=50	1.098609	-1.06006	1.12113	1.042217	1.090476
	n=100	1.062593	-0.99428	1.017271	0.988661	1.025071
	n=250	1.030955	-0.98289	0.97916	0.974852	0.977667
30% Missingness	n=50	1.104232	-0.94469	0.951333	0.965304	1.058915
	n=100	1.061597	-0.92397	0.956362	0.890659	0.958509
	n=250	1.034481	-0.92603	0.908687	0.879251	0.930022

Table 6. Parameter Estimates by Conditional MLE for Bayesian (MICE) Imputed Panel Data Set.

		Bayesian (MICE) Imputed Data				
		β1	β2	β3	β4	β5
10% Missingness	n=50	1.067832	-1.07822	1.27407	1.062909	1.141502
	n=100	1.028601	-1.00793	1.02702	1.010346	1.026907
	n=250	0.998693	-0.99862	0.994132	0.999532	1.000275
30% Missingness	n=50	0.966065	-1.04179	0.986995	0.99603	0.878896
	n=100	0.946108	-0.94759	0.95725	0.931001	0.834477
	n=250	0.917728	-0.93417	0.934447	0.921759	0.862901

Table 7. Comparative Parameter Biases and Determinants of the Hessian Matrices across Different Imputed Panel Data Sets with Varying sample sizes.

		Determinant of Hessian ($\times 10^{-6}$)	Parameter Bias				
			β1	β2	β3	β4	β5
Complete Unimputed Data	*n=50*	10.33104	0.118411	0.069170	0.112667	0.122587	0.105870
	n=100	13.67689	0.057570	0.023855	0.074032	0.033003	0.052396
	n=250	15.00124	0.027265	0.014057	0.008524	0.015673	-0.006792

Mean Imputed Data	*n=50*	2.509932	0.043838	-0.033998	-0.031738	0.008104	0.329941
	n=100	3.186349	0.004478	-0.066237	-0.029276	-0.070154	0.262566
	n=250	4.677663	-0.019858	-0.061184	-0.077084	-0.086455	0.238532
LVCF Imputed Data	*n=50*	11.67492	-0.317889	-0.392132	-0.373402	-0.334978	-0.032867
	n=100	18.52447	-0.339128	-0.391447	-0.379594	-0.395006	-0.074370
	n=250	20.43683	-0.366831	-0.408155	-0.405048	-0.407119	-0.088892
Median Imputed Data	*n=50*	4.156453	0.104232	-0.055310	-0.048667	-0.034696	0.058915
	n=100	4.049792	0.061597	-0.076035	-0.043638	-0.109342	-0.041491
	n=250	6.079974	0.034481	-0.073969	-0.091313	-0.120749	-0.069978
Bayesian (MICE) Imputed Data	*n=50*	1.510900	-0.033935	0.041789	-0.013006	-0.003970	-0.121104
	n=100	1.968110	-0.053892	-0.052413	-0.042750	-0.068999	-0.165523
	n=250	3.926050	-0.082272	-0.065831	-0.065554	-0.078241	-0.137099

5. Dcussion, Conclusion and Recommendation

Through the variation of the simulated sample sizes, the study confirms the asymptomatic properties of parameter bias. The results show that the parameter estimates improve with increasing sample size. The precision of the estimates asymptotically increase thereby making them more statistically significant.

The key objectives of this study was to focus on a method used to modify the conditional likelihood function through partitioning of the covariate matrix in a bid to curb the incidental parameter problem and to assess the susceptibility of the Hessian matrix of the log likelihood function on the imputation techniques employed in completing a panel data set with missing covariates.

Undeniably, of all the classical imputation techniques, mean and median imputation do not introduce much undue bias into the data set and therefore perform relatively better than last value carried forward technique and mode imputation. However, model based technique for imputation like MICE yields even better estimates with even more reduced bias and precision [20]. Figure 1 shows the varying and reducing trends of the parameter estimates across sample sizes and across imputation methods used in this study.
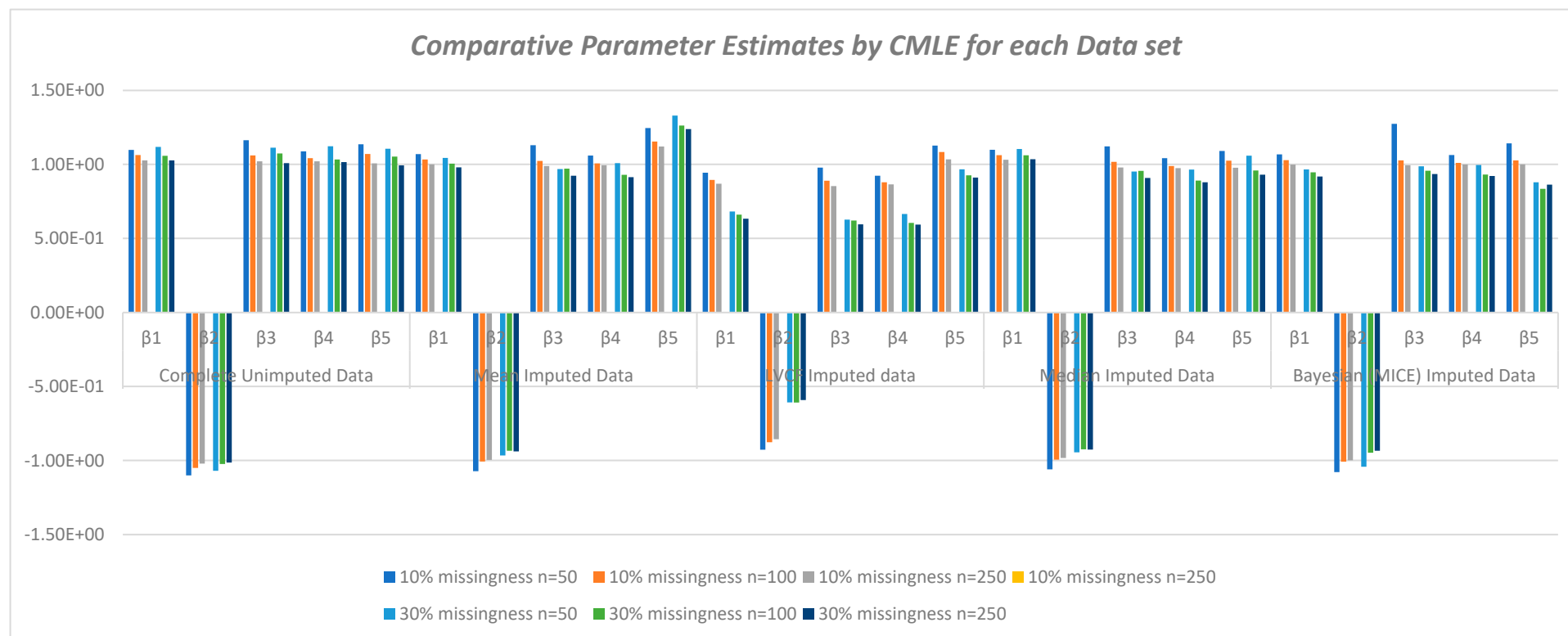
Figure 1. Comparative Parameter Estimates by Conditional MLE for Different Imputed Panel Data Sets with Varying sample sizes and Proportions of Missingness.
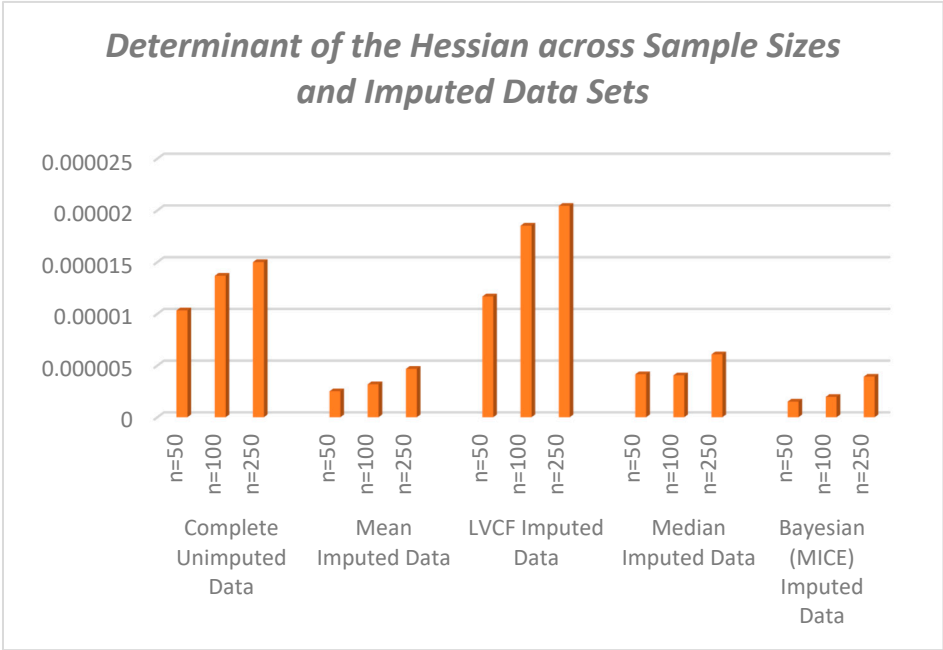
Figure 2. Comparative Determinants of the Hessian Matrices across Different Imputed Panel Data Sets with Varying sample sizes.

The value of Δx_{i_l} impacts inversely on the elements of the Hessian matrix and consequently its determinant. As seen, the study reveals that the smaller the determinant, the larger the parameter estimates which signifies an increased bias for smaller sample sizes. This indicates that by increasing the determinant of the hessian matrix through a reduction in Δx_{i_l} values, we reduce the product $J(\beta^{(h-1)})^{-1} s(\beta^{(h-1)})$ towards zero.

From the N-R algorithm (30), therefore, the inverse of the Hessian J serves to reduce the product $J(\beta^{(h-1)})^{-1} s(\beta^{(h-1)})$ to yield convergence in the iterations of $\beta^{(h)}$. An increasing Hessian modulus therefore ensures faster convergence of the parameter estimates with more precision as seen from table 7 and figure 2. The positive moduli of the Hessian for the conditional MLEs are sufficient for concavity of the log likelihood function that give the optimum estimates of the parameters.

A key importance of deriving the estimators is to increase the theoretical understanding of the estimators and also reduce the computational complexity while estimating logit panel models. As observed from the Monte Carlo results, unbalancedness in a data set biases the parameter estimates and the different imputation techniques employed in this study respond differently to the concavity of the Hessian matrix, hence the bias and efficiency of the estimates are affected too.

Although studies show that the within estimator performs relatively well in many standard situations, this study has demonstrated the advantage of working with the conditional likelihood function over the unconditional likelihood as a working trick to eliminate the fixed effects from the estimation process thereby limiting our concentration on the parameter estimates only.

As a recommendation, further developments can be done on this study by considering multiple fixed effects and observed time periods greater than T=2.

References

1. Janssen KJ, Donders ART, Harrell FE Jr., Vergouwe Y, Chen Q, Grobbee DE, Moons KG. Missing covariate data in medical research: to impute is better than to ignore. *Journal of Clinical Epidemiology* 2010; **63**(7):721–727.
2. Donders ART, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology* 2006; **59**(10):1087–1091.
3. Knol MJ, Janssen KJ, Donders ART, Egberts AC, Heerdink ER, Grobbee DE, Moons KG, Geerlings MI. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *Journal of Clinical Epidemiology* 2010; **63**(7):728–736.
4. Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics. Chapman and Hall, London.

5.   Firth, D. (1993) Bias Reduction of Maximum Likelihood Estimates. Biometrika, 80, 27-38. https://doi.org/10.1093/biomet/80.1.27

6.   Anderson, J.A. and Richardson, C. (1979) Logistic Discrimination and Bias Correction in Maximum Likelihood Estimation. Technometrics, 21, 71-78.

7.   McCullagh, P. (1986) The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data. Journal of the American Statistical Association, 81, 104-107. https://doi.org/10.1080/01621459.1986.10478244

8.   Shenton, L.R. and Bowman, K.O. (1977) Maximum Likelihood Estimation in Small Samples. Griffin's Statistical Monograph No. 38, London.

9.   S. Lee, "Detecting differential item functioning using the logistic regression procedure in small samples," *Applied Psychological Measurement*, vol. 41, no. 1, pp. 30–43, 2017.

10.   R. Puhr, G. Heinze, M. Nold et al., "Firth's logistic regression with rare events: accurate effect estimates and predictions?" *Statistics in Medicine*, vol. 36, no. 14, pp. 2302–2317, 2017.

11.   D. Firth, "Bias reduction of maximum likelihood estimates," *Biometrika*, vol. 80, no. 1, pp. 27–38, 1993.

12.   Chamberlain, G., (1980). Analysis of Covariance with Qualitative Data, *Review of Economic Studies* **47**, 225-238

13.   Matyas, L. and Lovrics, L. (1991). Missing observations and panel data- A Monte-Carlo analysis, *Economics Letters* **37**(1), 39-44.

14.   Neyman, J. and Scott, E.L., (1948). Consistent estimation from partially consistent observations. *Econometrica* **16**, 1-32.

15.   Lancaster, T., (2000). The incidental parameter problem since 1948, *Journal of Econometrics*, **95**, 391–413.

16.   Baltagi, B.H. (2001). *Econometric Analysis of Panel Data, 2nd edition*, New York, John Wiley.

17.   Hsiao, C., (2003). *Analysis of Panel Data*, 2nd edition, New York, Cambridge University Press.

18.   Greene, W., (2004a). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal* **7**(1): 98.

19.   Opeyo P.O., Olubusoye O.E. and Odongo L.O. (2014), *Conditional Maximum Likelihood Estimation for Logit Panel Models with Non-Responses*, International Journal of Science and Research, 3(7), 2242-2254

20.   Opeyo, P. , Cheng, W. and Xu, Z. (2023) Superiority of Bayesian Imputation to Mice in Logit Panel Data Models. *Open Journal of Statistics*, **13**, 316-358. doi: 10.4236/ojs.2023.133017.