International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Conditional Maximum Likelihood Estimation for Logit Panel Models with Non-Responses

Opeyo P.O.[1], Olubusoye O.E.[2], Leo O.Odongo[3]

[1]Pan African University, Institute of Basic Sciences, Technology and Innovation,
P. O. Box 62000- 00200 Nairobi, Kenya

[2]Department of Statistics University of Ibadan, Oyo-State, Nigeria

[3]Kenyatta University, Department of Mathematics,
P.O. Box 43844-00100, Nairobi, Kenya

Abstract: ***In analyzing most survey data in which the dependent variable is a binary choice variable taking values 1 or 0 for success or failure respectively it is feasible to consider the conditional probabilities of the dependent variable. Under strict exogeneity, this conditional probability equals the expected value of the dependent variable. This treatment calls for a nonlinear function which will ensure that the conditional probability lies between 0 and 1 and such functions yield the probit model and the logit model. For panel data econometrics, such nonlinear panel models require conditioning the probabilities on the minimum sufficient statistic for the fixed effects so as to curb the incidental parameter problem. Solving the joint p.d.f by maximum likelihood method yields consistent 'conditional maximum likelihood estimate' for the model parameters in cases when the data set is complete (or balanced) with no cases of missing observations. In cases of missing observations in the covariates, researchers employ several imputation techniques are used to make the data complete. Imputation, however, brings about a bias in the covariate and this bias is propagated to the parameter estimates. This study considers the susceptibility of nonlinear logit panel data model with single fixed effects to imputation by investigating the bias arising from various imputation methods. The study developed a conditional maximum likelihood estimator for nonlinear binary choice logit panel model in the presence of missing observations. A Monte Carlo simulation was designed to determine the magnitude of bias arising from common imputation techniques and recommend better techniques to be used in order to improve model performance in the presence of missing observations in econometrics panel data analysis. The simulation results show that the conditional logit estimator presented in this paper is less biased than the unconditional logit estimators without sacrificing on the precision.***

Keywords: Panel Data, Binary choice, Imputation, Monte Carlo, Bias, Conditional Maximum Likelihood

1. Introduction

1.1 Background Introduction

A general panel data model is of the form

$$y_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + c_i + \gamma_t + u_{it}, i = 1,\ldots,N;\ t = 1,\ldots,T \quad (1)$$

Where the parameters c_i and γ_t represent the individual specific and time specific effects respectively. Assuming only the individual specific effects c_i then the equation (1) takes the form

$$y_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + c_i + u_{it} \quad (2)$$

The relationship between c_i and u_{it} determines whether the relation (2) is treated as fixed or random effects model. This is to say that if c_i is correlated with $\mathbf{x_{it}}$ then the model has only u_{it} as the stochastic part and c_i is treated as fixed (non-random). As such, we have a fixed effect panel model. On the other hand, it is a random effect model, if it becomes part of the stochastic part of (2) so that

$$y_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + v_{it} \quad (3)$$

Where $v_{it} = c_i + u_{it}$. Equation (3) is the random effect model.

In estimating panel model parameters, therefore, there exist generally two categories of models, fixed effects models (FE) and random effects (RE) models. With the former, one does not estimate the effects of the variables that are individual specific and time invariant but rather controls for them or 'partials them out'. The later (RE models) estimate the effects of these time invariant variables. These estimates may be biased since other omitted variables are uncontrolled for.

If the dependent variable y_{it} is continuous then the parameters in panel data model can be estimated. The approaches used so far in estimating panel models with fixed effects aim at controlling for these effects by eliminating the presence of these effects from the model and estimate the coefficients of the regressors. If on the other hand the dependent variable is categorical, then specific nonlinear functions that preserve the structure of the dependent variable are considered. Such nonlinear functions include among others, the logit, probit and poisson models. Among the approaches explored to estimate fixed effects models include:

- Demeaning variables- where the within subject means (averages) are subtracted from each observed value of the variables thereby eliminating the constant nuisance factor for each subject. This approach is known to work best for linear regression models but fails in logistic regression.
- Unconditional maximum likelihood - here dummy variables are created for each subject (except one) and included in the model i.e. N-1 dummies introduced. Estimating linear regressions by unconditional maximum likelihood produces consistent estimates with the demeaning variables method but for logistic regressions, these estimates are biased.
- Conditional maximum likelihood estimation – this is the most preferred method for logistic regressions. Here, the

conditional maximum likelihood 'conditions' the (fixed effects) out of the likelihood function [7]. This is done by conditioning the likelihood function on the total number of events observed for each subject.

The concepts of conditional maximum likelihood for nonlinear panel models has been tackled in several studies from cases with only a single fixed effect to multiple fixed effects. For static linear models, consistent estimates for the parameters are obtained by simply differencing out the fixed effects. For nonlinear panel models, however, there exist the well-known incidental parameter problem realized by Neyman and Scott [24] in which the number of fixed effects increases with increasing sample size. Incidental parameters are such parameters whose dimension increases with sample size. For example, as N approaches infinity, the number of fixed effects increases and so they are incidental parameters. Such parameters cannot be consistently estimated [5]. Other attempts to solve the incidental parameter problem succeeded for the Poisson and negative binomial models with single fixed effects [15]. Manski [21] generalized the logit model and developed a conditional maximum score estimation of binary response models.

Charbonneau [9] developed the works of Hausman, Hall and Griliches [15] by considering the adaptability of nonlinear panel models to multiple fixed effects. From Monte Carlo simulations [9], the conditional ML logit estimator proved less biased than other logit estimators. As much parameter estimation of panel models is possible, complications arise when the panels are unbalanced. Such unbalancedness in panel data is brought about by delayed entry, early exit or intermittent non-response from a study unit. For the former two causes of unbalancedness, each individual is observed $T_i$times and analysis of the panel models is still feasible. However, in cases of intermittent non-responses a need to establish the nature and cause of the non-response suffices. Approaches suggested in literature on how to handle missing observations become valid in such cases. Moreover, to avoid serious inferential problems which may arise from sample nonresponses thereby misdirecting policy actions, much attention need to be given to the problem of non-response bias both at stages of data collection and data analysis. This study therefore examines the impact of missing data to the conditional maximum likelihood estimation procedures of nonlinear panel models for discrete choice dependent variable. We derive a conditional maximum likelihood estimator with reduced bias for nonlinear binary response logit panel models in the presence of imputed missing observations. Using simulations with various types of missing data to evaluate the magnitude of bias arising from using common imputation techniques to estimate missing observations, we shall attempt to recommend the best techniques to be used in order to improve the treatment of missing data.

2. The Review of Literature

Panel data econometrics has greatly developed since the handbook chapter by Chamberlain [8]. Panel data methods so far studied are necessary for understanding individual specific behaviors. The analysis of two way models, both fixed and random effects, has been well worked out in the linear case in studies by Baltagi[4], [5]. Greene [12] shows that individual specific dummy variable coefficients can be estimated using group specific averages of residuals. By least squares dummy variables (LSDV) approach, the slope parameters in linear models can also be estimated using simple first differences.

Although for linear cases, regression using mean deviations sweeps out the fixed effects, there are a few analogous cases of nonlinear models that have been identified in literature. Among them are the binomial logit model [12], Poisson and negative binomial regressions [15] and exponential regression model [13], [23]. Differently put, when studying static linear models, fixed effects do not generally cause any problem, since they can easily be differenced out to allow consistent estimation of the relevant parameters. However, when considering nonlinear panel data models, the incidental parameter problem identified by Neyman and Scott [24], motivated a rich literature on the estimation of single fixed effects nonlinear panel data models. Rasch considered the first model in the literature - the logit model [26], [27]. Later, Manski[21] generalized this to develop a conditional maximum score estimator for binary response models that remains consistent under weak assumptions on the distribution of the errors. On the same breath, Hausman, Hall and Griliches used the relationship between the Poisson and multinomial distribution to solve the incidental parameter problem in the Poisson regression model (and Negative Binomial) in the presence of a single fixed effect [15]. Like in the logit case, this results in a conditional likelihood approach that can be used to consistently estimate the parameters of interest.

With a more general approach to the problem, Hahn and Newey [14] show that when N and T grow at the same rate, the fixed effects estimator is asymptotically biased and the asymptotic confidence intervals are wrong. They suggest two bias correction methods (the panel Jackknife and the analytic bias correction).

Most of these models are however considered majorly for cases with balanced panels in which no missing data due to nonresponses exist. The problem of non-response is normally ignorable for a regression model of interest if inference can be made about the model without caring about the process that causes the missing data. Certain conditions that allow one to neglect the selection process are given by Rubin for cross sectional case [19], [30]. Specifically, these authors introduced the concepts of missing at random (MAR), missing not at random (MNAR) and missing completely at random (MCAR). Nikos Tsikriktis [25] gave detailed overviews on various techniques of dealing with missing data which he categorized into three: deletion procedures, replacement procedures and model based procedures.

Griliches and Hausman [15] note that a frequent drawback of using panel data is the insignificant results produced by the 'within' approach to their analysis, which are often blamed on the errors of measurement magnified by this approach. They provide a variety of errors-in-variables models for panel data, but for a continuous dependent variable. When the dependent variable is discrete, the problem changes.

Stefanski and Carroll (1985) study errors in variables in the logistic regression model and suggest a bias-adjusted estimator. Kao and Schnell (1987) extend the results to panel data and show that, with errors in variables, the conditional maximum-likelihood estimator for a binary regression model for a panel is asymptotically biased. They also introduce a bias-corrected estimator, which is examined asymptotically when the measurement error is small but non-negligible.

Individuals present in the data base may not be observed during the same period (unbalanced panels) or there may be 'holes' in the observation panel leading to incomplete panels. In literature, there exist two possibilities of estimating an econometric model with these kinds of incomplete panels. We can either use appropriate (unbalanced) estimation methods [3], [6], which are in general quite complex or drop from the panel those individuals for which the observations are not complete and carry out the estimation on a balanced and complete sub-panel of the original one.

Verbeek-Nijman (1990) show that if we have unbalanced or incomplete panels and we use the usual estimators of panel models based on a balanced and complete sub-panel, these are (asymptotically) unbiased and consistent (except the OLS) under quite general and reasonable conditions in the case when the observations are missing at random (so there is no selectivity bias present).

From the available literature, it is evinced that not much study on the panel data econometrics for the logit model has explored the concepts of nonresponse bias. As much as imputation techniques exist that can make datasets complete for ease of parameter estimation, the magnitudes of the biases induced into the parameter estimates are not substantively quantified. This means that there does not exist concrete procedures that can be used to pick on the best imputation technique in the estimation of panel models. A study in this line will therefore add on to the existing theoretical knowledge. The available literature thus iterates that conditional maximum likelihood estimates are consistent even for the logit model although with smaller bias compared to the unconditional MLE. Imputation also biases the covariates' averages. A study that combines these two biases, due to logit regression and imputation, is worthwhile.

3. Materials and Methods

3.1 Binary Choice Panel Models

3.1.1 Binary Choice Variable

In many economic studies, the dependent variable is categorical indicating a success or a failure of an event. Such dependent variable is normally represented by a binary choice variable $y_{it} = 1$ if the event happens and 0 if it does not happen for individual i at time t. In fact if p_{it} is the probability of success for individual i at time t, then $E(y_{it}) = 1 \times p_{it} + 0 \times (1 - p_{it}) = p_{it}$ and this is usually modeled as a function of some explanatory variables

$$p_{it} = Pr(y_{it} = 1) = E(y_{it}|x_{it}, c_i) = F(\boldsymbol{x_{it}\beta} + c_i)$$

We first consider the linear regression model $y_{it} = \boldsymbol{x_{it}\beta} + c_i + u_{it}$ (4) where y is a binary response variable, $\boldsymbol{x_{it}}$is a 1xK vector of observed explanatory variables (including a constant), $\boldsymbol{\beta}$ is a K x1 vector of parameters, $c_i$is an unobserved time invariant individual effect, and u_{it} is a zero-mean residual uncorrelated with all the terms on the right-hand side. Here, we assume strict exogeneity holds i.e. the residual $u_{it}$is uncorrelated with all x-variables over the entire time period spanned by the panel.

Since the dependent variable is binary, it is natural to interpret the expected value of y as a probability. Indeed, under random sampling, the unconditional probability that y equals one is equal to the unconditional expected value of y,

i.e. E (y) = Pr (y = 1). As such, $Pr(y_{it} = 1|\boldsymbol{x_{it}}, c_i) = E(y_{it} = 1|\boldsymbol{x_{it}}, c_i; \beta)$

So if the model (4) above is correctly specified, we have

$$\left.\begin{aligned} Pr(y_{it} = 1|\boldsymbol{x_{it}}, c_i) &= \boldsymbol{x_{it}\beta} + c_i \\ Pr(y_{it} = 0|\boldsymbol{x_{it}}, c_i) &= 1 - (\boldsymbol{x_{it}\beta} + c_i) \end{aligned}\right\} \quad (5)$$

Equation (5) is a binary response model. In this particular model the probability of success (i.e. y = 1) is a linear function of the explanatory variables in the vector x. Hence this is called a linear probability model (LPM) which can be used to estimate the parameters, such as OLS or the within estimator. This LPM however has limitations when used to estimate the parameters for a discrete choice variable. One undesirable property of the LPM, among others, is that we can get predicted "probabilities" either less than zero or greater than one. Of course a probability by definition falls within the (0, 1) interval, so predictions outside this range are meaningless and somewhat embarrassing.

To address the problems of LPM, a nonlinear binary response model is used where we write our nonlinear binary response model as

$$Pr(y_{it} = 1|\boldsymbol{x_{it}}, c_i) = G(\boldsymbol{x_{it}\beta} + c_i) \quad (6)$$

with G being a function taking on values strictly between zero and one: i.e.0 < G(z) < 1, for all real numbers z. The fact that$0 < G(\boldsymbol{x_{it}\beta} + c_i) < 1)$ ensures that the estimated response probabilities are strictly between zero and one, which thus addresses the main limitation of using LPM. G is a cumulative density function (cdf), monotonically increasing in the index z (i.e. $\boldsymbol{x_{it}\beta} + c_i$), with

$$\left.\begin{aligned} Pr(y_{it} = 1|\boldsymbol{x_{it}}, c_i) \to 1 \; as \; \boldsymbol{x_{it}\beta} + c_i \to \infty \\ Pr(y_{it} = 1|\boldsymbol{x_{it}}, c_i) \to 0 \; as \; \boldsymbol{x_{it}\beta} + c_i \to -\infty \end{aligned}\right\} \quad (7)$$

Thus G is a nonlinear function, and hence we cannot use a linear regression model for estimation. Various non-linear functions for G have been suggested in the literature and the most common ones are the logistic distribution, yielding the logit model, and the standard normal distribution, yielding the probit model. In the logit model, G takes the form,

$$G(\boldsymbol{x_{it}\beta} + c_i) = \frac{e^{x_{it}\beta + c_i}}{1 + e^{x_{it}\beta + c_i}} \quad (8)$$

which is between zero and one for all values of $\boldsymbol{x_{it}\beta}$. This is the cumulative distribution function (CDF) for a logistic variable.

3.3.2 Assumptions of the Logit Model

In Logit estimation, there does not exist many of the key assumptions of linear regression and general linear models that are based on ordinary least squares algorithms – particularly regarding linearity, normality, homoscedasticity, and measurement level. As such, logit regression has certain unique characteristics to be mentioned: (1) it does not need a linear relationship between the dependent and independent variables. Logistic regression can handle all sorts of relationships, because it applies a non-linear log transformation to the predicted odds ratio,(2) the independent variables do not need to be multivariate normal – although multivariate normality yields a more stable solution. Also the error terms (the residuals) do not need to be multivariate normally distributed, (3) homoscedasticity is not needed, (4) it can handle ordinal and nominal data as independent variables. The independent variables do not need to be metric (interval or ratio scaled).

3.4 Estimation of Logit Model

3.4.1 Incidental parameter Problem

For Panel data, the presence of individual effect complicates the parameter estimation significantly. Consider the fixed effects panel data model

$y_{it} = \boldsymbol{x_{it}\beta} + c_i + u_{it}$ with $Pr(y_{it} = 1) = F(\boldsymbol{x_{it}\beta} + c_i)$.

In this case 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 c_i cannot be consistently estimated for fixed T. This is known as the incidental parameter problem in statistics, first discussed by Neyman and Scott (1948) and later reviewed by Lancaster (2000).

For linear panel data regression model, when T is fixed, 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 (Hsiao 2003). For qualitative binary choice model with fixed T, this is not possible as demonstrated by Chambelain (1980).

Hsiao (2003) simply illustrates how the inconsistency of the MLE of c_i is transmitted into inconsistency for $\widehat{\boldsymbol{\beta}}_{mle}$. This is done in the context of a logit 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 (2004a) 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. Using Monte Carlo experiments, he shows that the fixed effects MLE is biased even when T is large. For N = 1000, T = 2 and 200 replications, this bias is 100%, confirming the results derived by Hsiao (2003). However, this bias improves as T increases. For example, when N = 1000 and T = 10 this bias is 16% and when N = 1000 and T = 20 this bias is 6.9%.

3.4.2 The Unconditional likelihood function

The logit model is estimated by means of Maximum Likelihood (ML). That is, the ML estimate of β is the particular vector $\hat{\beta}^{ML}$ that gives the greatest likelihood of observing the outcomes in the sample $\{y_1, y_2, \ldots\}$ conditional on the explanatory variables x.

By assumption, the probability of observing y_{it}= 1 is G$(\boldsymbol{x_{it}\beta} + c_i)$ while the probability of observing y_{it}= 0 is $1 - G(\boldsymbol{x_{it}\beta} + c_i)$: It follows that the probability of observing the entire sample is

$$L(y|\boldsymbol{x};\beta) = \prod_{i \in l} G(\boldsymbol{x_{it}\beta} + c_i) \prod_{i \in m} [1 - G(\boldsymbol{x_{it}\beta} + c_i)] \quad (9)$$

Where l refers to the observations for which y = 1 and m to the observations for which y = 0.

We can rewrite this as

$$L(y|\boldsymbol{x};\beta) = \prod_{i=1}^{N} (G(\boldsymbol{x_{it}\beta} + c_i))^{y_i} [1 - G(\boldsymbol{x_{it}\beta} + c_i)]^{1-y_i} \quad (10)$$

because when y = 1 we get $G(\boldsymbol{x_{it}\beta} + c_i)$ and when y = 0 we get $[1 - G(\boldsymbol{x_{it}\beta} + c_i)]$.

The log likelihood for the sample is

$$lnL(y|\boldsymbol{x};\beta) = \sum_{i=1}^{N} \{y_i lnG(\boldsymbol{x_{it}\beta} + c_i) + (1 - y_i) ln[1 - G(\boldsymbol{x_{it}\beta} + c_i)]\} \quad (11)$$

The MLE of β maximizes this log likelihood function.

3.4.3 Conditional Likelihood function for Logit Panel Model

If G is the logistic CDF then we obtain the logit log likelihood:

$$lnL(y|\boldsymbol{x};\beta) = \sum_{i=1}^{N} \left\{ y_i ln\left(\frac{e^{x_{it}\beta + c_i}}{1 + e^{x_{it}\beta + c_i}}\right) + (1 - y_i) ln\left(\frac{1}{1 + e^{x_{it}\beta + c_i}}\right) \right\} \quad (12)$$

Estimating the parameters in this model is not easy as it is specified since the unobserved individual characteristics, c_i are also not known. In linear models, it is easy to eliminate c_i by means of first differencing or using within transformation. If we attempt to estimate c_i directly by adding N-1 individual dummy variables to the logit specification, this will result in severely biased and inconsistent estimates of β unless T is large due to the incidental parameters problem.

One important advantage of the logit model over the probit model is that it is possible to obtain a consistent estimator of β without making any assumptions about how c_i is related to $\boldsymbol{x_{it}}$ (however, strict exogeneity must hold).

This is possible, because the logit functional form enables us to eliminate c_i from the estimating equation, once we condition on the "minimum sufficient statistic" for c_i. As such we obtain the conditional likelihood function whose parameters are estimated. For T = 2, the conditional probabilities:

$Pr(y_{i1} = 0, y_{i2} = 1|x_{i1}, x_{i2}, c_i, y_{i1} + y_{i2} = 1)$ $(13a)$ and

$Pr(y_{i1} = 1, y_{i2} = 0|x_{i1}, x_{i2}, c_i, y_{i1} + y_{i2} = 1)$ $(13b)$ are expressed as;

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

$$Pr(y_{i1} = 1, y_{i2} = 0|x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1) = \frac{1}{1 + e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}} \quad (17)$$

It also follows that probabilities (16) and (17) are conditional on $y_{i1} + y_{i2} = 1$and are independent of c_i.

The distribution function is thus given as

$$Pr(y_{i1}, y_{i2}|x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1) = \begin{cases} 1 \; if \; (y_{i1}, y_{i2}) = (0,0) or (1,1) \\ \frac{1}{1+e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}} \; if \; (y_{i1}, y_{i2}) = (1,0) \\ \frac{e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}}{1+e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}} \; if \; (y_{i1}, y_{i2}) = (0,1) \end{cases} \quad (18)$$

Hence, by maximizing the following conditional log likelihood function

$$lnL = \sum_{i=1}^{N} \left\{ d_{01i} ln \left(\frac{e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}}{1 + e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}} \right) + d_{10i} ln \left(\frac{1}{1 + e^{(x_{i2}-x_{i1})\boldsymbol{\beta}}} \right) \right\} \quad (19)$$

we obtain consistent estimates of β, regardless of whether c_i and x_{it} are correlated.

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

$$Pr\left(y_{i1}, y_{i2}, \ldots, y_{iT} \middle| \boldsymbol{X_i}, \sum_t y_{it} \right) = \frac{e^{(\sum_t y_{it} x_{it} \boldsymbol{\beta})}}{\sum_{d \in B_i} e^{(\sum_t d_{it} x_{it} \boldsymbol{\beta})}} \quad (20)$$

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

4. Methodology and Data Analysis

4.1 Parameter Estimation

Consider the logit panel data model given $P(\boldsymbol{y_{it}} = \boldsymbol{1}|\boldsymbol{x_{it}}, \boldsymbol{\beta}, c_i) = \frac{e^{x_{it}\beta + c_i}}{1+e^{x_{it}\beta + c_i}}$ (22) where $\boldsymbol{x_{it}}$is the vector of covariates. In the presence of missing observations in the 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, the equations (16) and (17) are expressible as

$$Pr(y_{i1} = 0, y_{i2} = 1|\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, y_{i1} + y_{i2} = 1) = \frac{e^{\{(x_{i2_s}+x_{i2_I})-(x_{i1_s}+x_{i1_I})\}\boldsymbol{\beta}}}{1 + e^{\{(x_{i2_s}+x_{i2_I})-(x_{i1_s}+x_{i1_I})\}\boldsymbol{\beta}}} \quad (27)$$

$$Pr(y_{i1} = 1, y_{i2} = 0|\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, y_{i1} + y_{i2} = 1) = \frac{1}{1 + e^{\{(x_{i2_s}+x_{i2_I})-(x_{i1_s}+x_{i1_I})\}\boldsymbol{\beta}}} \quad (28)$$

Equations (27) and (28) can as well be expressed as

$$Pr(y_{i1} = 0, y_{i2} = 1|\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, y_{i1} + y_{i2} = 1) = \frac{e^{\Delta x_{i_I}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_I}\boldsymbol{\beta}}} \quad (29)$$

$$Pr(y_{i1} = 1, y_{i2} = 0|\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, y_{i1} + y_{i2} = 1) = \frac{e^{-\Delta x_{i_s}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_I}\boldsymbol{\beta}}} \quad (30)$$

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

The conditional log likelihood function can thus be obtained using equations (29) and (30) as

$$lnL = \sum_{i=1}^{N} \left\{ d_{01i} ln \left(\frac{e^{\Delta x_{i_I}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_I}\boldsymbol{\beta}}} \right) + d_{10i} ln \left(\frac{e^{-\Delta x_{i_s}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_I}\boldsymbol{\beta}}} \right) \right\} \quad (31)$$

4.2 Newton-Raphson Algorithm

Maximization of equation (31) can be performed by the Newton-Raphson algorithm. Starting from an initial estimate $\boldsymbol{\beta}^{(0)}$, the algorithm consists of iterating the estimate at step h as

$$\boldsymbol{\beta}^{(h)} = \boldsymbol{\beta}^{(h-1)} + \boldsymbol{J}\left(\boldsymbol{\beta}^{(h-1)}\right)^{-1} \boldsymbol{s}\left(\boldsymbol{\beta}^{(h-1)}\right) \; \mathbf{(32)}$$

Where, $\boldsymbol{s}(\boldsymbol{\beta}) = \frac{\partial lnL}{\partial \boldsymbol{\beta}}$is the score vector and $\boldsymbol{J}(\boldsymbol{\beta}) = -\frac{\partial^2 lnL}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}$is the observed information matrix given respectively as

$$s(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{lnL}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left\{ d_{01i} \left[\Delta \boldsymbol{x'}_{i_l} - \left(\frac{-\Delta \boldsymbol{x'}_{i_s} e^{-\Delta x_{i_s}\boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_l} e^{\Delta x_{i_l}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_l}\boldsymbol{\beta}}} \right) \right] - d_{10i} \left[\Delta \boldsymbol{x}_{i_s} + \left(\frac{-\Delta \boldsymbol{x'}_{i_s} e^{-\Delta x_{i_s}\boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_l} e^{\Delta x_{i_l}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_l}\boldsymbol{\beta}}} \right) \right] \right\} \quad (32)$$

$$\boldsymbol{J}(\boldsymbol{\beta}) = -\frac{\partial^2 \boldsymbol{lnL}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta'}} = \sum_{i=1}^{N} \left\{ d_{01i} \left[\left(\frac{\Delta \boldsymbol{x}_{i_s}^{\,2} e^{-\Delta x_{i_s}\boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_l}^{\,2} e^{\Delta x_{i_l}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_l}\boldsymbol{\beta}}} \right) - \left(\frac{-\Delta \boldsymbol{x'}_{i_s} e^{-\Delta x_{i_s}\boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_l} e^{\Delta x_{i_l}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_l}\boldsymbol{\beta}}} \right)^2 \right] + d_{10i} \left[\left(\frac{\Delta \boldsymbol{x}_{i_s}^{\,2} e^{-\Delta x_{i_s}\boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_l}^{\,2} e^{\Delta x_{i_l}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_l}\boldsymbol{\beta}}} \right) - \left(\frac{-\Delta \boldsymbol{x'}_{i_s} e^{-\Delta x_{i_s}\boldsymbol{\beta}} + \Delta \boldsymbol{x'}_{i_l} e^{\Delta x_{i_l}\boldsymbol{\beta}}}{e^{-\Delta x_{i_s}\boldsymbol{\beta}} + e^{\Delta x_{i_l}\boldsymbol{\beta}}} \right)^2 \right] \right\}$$

4.3 Monte Carlo Simulation

In this section, we present Monte Carlo evidence to support the conditional ML fixed effect estimator developed above. For this, we focus on the logit estimator given by the maximization of equation (31). The simulations will compare that estimator to the unconditional logit estimator which estimates all the fixed effects by putting in dummies. The unconditional logit estimator, however, is subject to the incidental parameter problem.

To account for different possible features of the data, this comparison will be made for two sets of data, one complete (balanced) and the other incomplete (unbalanced) due to intermittent nonresponses. The latter data set is balanced by imputing the missing observations and substituting the imputed vector $\boldsymbol{x}_{it_l}$into the conditional log likelihood function (31) where the imputation methods described in section 3.7 are employed. Both panel sets are applied to the estimation of the following model:

$y_{it} = 1(\boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \geq 0)\ i = 1,2,\dots,n\ t = 1,2$ where $\boldsymbol{x}_{it}$ is a vector of five explanatory variables drawn from uniform, binomial and normal distributions and the error term u_{it} is drawn from a normal distribution. The variables' descriptions areas in table 1. All other parameters, beta1 to beta5, of the model necessary to calculate the dependent variable y were fixed as β1=1, β2=-1, β3=1, β4=1 and β5=1. Having determined these variables, the dependent variable, y, was calculated from the relation $y_{it} = 1\left(c_i + \beta_1 x^1{}_{it} + \beta_2 x^2{}_{it} + \beta_3 x^3{}_{it} + \beta_4 x^4{}_{it} + \beta_5 x^5{}_{it} x_5 + v_{it} \geq 0\right)\ i = 1,2,\dots,n\ t = 1,2,\dots,T$ where v_{it} is a logistic variable given by $v_{it} = ln\left|\frac{u_{it}}{1+u_{it}}\right|$ with u_{it} being a standard normal random variable. The fixed effects $c_i$are obtained as functions of x1 and t by the relation $c_i = \frac{\sqrt{t}\sum x1}{n} + a_i$with a_i being a standard normal random variable as well.

Table 1: Description of variables

Variable	*Type*	
x1	continuous	N~(0, 1)
x2	continuous	U~ (0, 1)
x3	continuous	N~ (0.5, 0.5)
x4	discrete	B~(nT, 2, 0.65)
x5	discrete	binomial

Three different sample sizes were used for both sets of data estimated i.e. n = 50, 100 and 250.In addition, for each sample size; we vary the proportion of missingness from 10% to 30% by randomly deleting the desired proportion of observations from the data set and imputing them back through mean imputation, last value carried forward imputation and median imputation. Whenever fixed effects are estimated, the coefficients are truncated in order to ensure convergence. The summarized results for 1000 replications are given in Tables 2 to 7. For both estimators (unconditional logit and conditional logit) considered, we report the median bias, the median absolute deviation (MAD), the mean bias, and the root mean squared error (RMSE) for all the four coefficient estimates.

Table 2: Sample size n= 50, T=2, percentage of missingness=10%

Model	*Balanced/Unbalanced*		*Parameter*	*Median Bias*	*MAD*	*Mean Bias*	*RMSE*
Logit (With FE)	Balanced		β_1	-0.09098714	0.2961396	-0.1288857	0.4743939
			β_2	0.17175351	0.7306155	0.1752378	1.1298963
			β_3	-0.06030911	0.4217924	-0.1165433	0.6785884
			β_4	-0.07580644	0.3290792	-0.1192750	0.5285040
			β_5	-0.13644092	0.5041175	-0.2033034	1.1341030
	Unbalanced (But imputed)	Mean Imputation	β_1	-0.05746778	0.2935424	-0.10273445	0.4628570
			β_2	0.11475929	0.7680350	0.12893202	1.1360577
			β_3	-0.03918777	0.4128462	-0.07500963	0.6733769
			β_4	-0.04251258	0.3367287	-0.08647667	0.5289352
			β_5	-0.23493028	0.4945531	-0.30105814	1.3779656
		Last Value carried forward	β_1	0.057110998	0.2434638	0.03976534	0.4067278
			β_2	-0.007469449	0.7141621	-0.01947585	1.0668631
			β_3	0.106552433	0.4020277	0.07252765	0.6392440
			β_4	0.091170987	0.3080340	0.06379989	0.4810950

			β_5	-0.157570806	0.4619799	-0.20836854	1.0964141
		Median Imputation	β_1	-0.09354977	0.2956615	-0.13402056	0.4717615
			β_2	0.09513642	0.7587241	0.11511149	1.1284732
			β_3	-0.02928728	0.4165188	-0.06855655	0.6738861
			β_4	-0.02013126	0.3355964	-0.06199504	0.5192618
			β_5	-0.11400131	0.4626130	-0.15238442	1.0552691
Conditional Logit	Balanced		β_1	-0.07615552	0.2935767	-0.1126502	0.4625236
			β_2	0.15074006	0.7218983	0.1585480	1.1108020
			β_3	-0.04510045	0.4168382	-0.1005548	0.6654922
			β_4	-0.05992669	0.3233472	-0.1029941	0.5162705
			β_5	-0.12230168	0.4948172	-0.1907195	1.1798448
	Unbalanced (But imputed)	Mean Imputation	β_1	-0.04350537	0.2888392	-0.08732180	0.4522001
			β_2	0.09972025	0.7565596	0.11338074	1.1181839
			β_3	-0.02588503	0.4070878	-0.06007568	0.6617154
			β_4	-0.02890030	0.3289778	-0.07115199	0.5181301
			β_5	-0.22017651	0.4863997	-0.28630585	1.4248434
		Last Value carried forward	β_1	0.07014586	0.2385731	0.05283088	0.4014827
			β_2	-0.02057578	0.7055523	-0.03258641	1.0521639
			β_3	0.11620920	0.3950074	0.08516777	0.6312142
			β_4	0.10417631	0.3039078	0.07672558	0.4753835
			β_5	-0.14369129	0.4552648	-0.19646296	1.1464525
		Median Imputation	β_1	-0.078249682	0.2903286	-0.11828839	0.4601677
			β_2	0.080372137	0.7480488	0.09986714	1.1110511
			β_3	-0.016243206	0.4123068	-0.05381635	0.6624871
			β_4	-0.007711298	0.3316286	-0.04715232	0.5094865
			β_5	-0.099931330	0.4572849	-0.14189014	1.1211302

Table 3: Sample size n= 50, T=2, percentage of missingness=30%

				Sample size n= 50, T=2, percentage of missingness=30%			
Model	*Balanced/Unbalanced*		*Parameter*	*Median Bias*	*MAD*	*Mean Bias*	*RMSE*
Logit (With FE)	Balanced		β_1	-0.09768892	0.2815223	-0.1323880	0.4512832
			β_2	0.08893815	0.6959374	0.1506918	1.1535639
			β_3	-0.13115775	0.4111136	-0.1439335	0.6980234
			β_4	-0.08391560	0.3190010	-0.1294278	0.5217967
			β_5	-0.11427075	0.4932405	-0.1848617	1.2574775
	Unbalanced (But imputed)	Mean Imputation	β_1	-0.04624262	0.2744833	-0.06820391	0.4571537
			β_2	0.06214297	0.7773881	0.06100301	1.2151814
			β_3	-0.01779544	0.4811433	-0.03848048	0.7640412
			β_4	0.02343789	0.3521519	-0.01490505	0.5469861
			β_5	-0.33329933	0.5109296	-0.36369550	0.8748120
		Last Value carried forward	β_1	0.30291813	0.2137996	0.28007086	0.4434043
			β_2	-0.38837327	0.6439265	-0.32954224	1.0662600
			β_3	0.34860522	0.3847189	0.32062928	0.7052962
			β_4	0.35828686	0.2717921	0.33350304	0.5500622
			β_5	0.01685191	0.4217159	-0.02104363	0.9962462
		Median Imputation	β_1	-0.081919079	0.2602768	-0.12294985	0.4630715
			β_2	0.009853557	0.7636018	0.04468553	1.2013292
			β_3	-0.013160454	0.4715510	-0.02060252	0.7567571
			β_4	0.078381729	0.3487885	0.03521928	0.5588983
			β_5	-0.009659650	0.4590397	-0.18470213	1.6086436
Conditional Logit	Balanced		β_1	-0.08147882	0.2781479	-0.1160803	0.4395600
			β_2	0.07675574	0.6865320	0.1340337	1.1341241
			β_3	-0.11678069	0.4051523	-0.1274709	0.6840850
			β_4	-0.06985921	0.3150342	-0.1129801	0.5093962
			β_5	-0.09913601	0.4863511	-0.1756863	1.3441573

	Unbalanced (But imputed)	Mean Imputation	β_1	-0.032836315	0.2711211	-0.054044888	0.4482092
			β_2	0.047383313	0.7655643	0.047190264	1.1978881
			β_3	-0.006073291	0.4762234	-0.024771599	0.7525678
			β_4	0.035336783	0.3469567	-0.001514772	0.5387079
			β_5	-0.317244908	0.5035051	-0.346235313	0.8575039
		Last Value carried forward	β_1	0.31154713	0.2106966	0.28907710	0.4453732
			β_2	-0.39472664	0.6366078	-0.33786347	1.0564189
			β_3	0.35598466	0.3805953	0.32913237	0.7018953
			β_4	0.36615453	0.2685745	0.34189440	0.5504764
			β_5	0.02824329	0.4165644	-0.01311922	1.0638477
		Median Imputation	β_1	-0.069542871	0.2534770	-0.108213966	0.4525874
			β_2	-0.002605846	0.7515572	0.031160962	1.1843682
			β_3	-0.000713831	0.4636996	-0.007204582	0.7457460
			β_4	0.089964923	0.3449030	0.047757450	0.5516115
			β_5	0.005607919	0.4508094	-0.187323187	1.7749189

Table 4: Sample size n= 100, T=2, percentage of missingness=10%

				Sample size n= 100, T=2 , percentage of missingness=10%			
Model	Balanced/Unbalanced		Parameter	Median Bias	MAD	Mean Bias	RMSE
Logit (With FE)	Balanced		β_1	-0.07228644	0.1842174	-0.07599035	0.2920723
			β_2	0.08822273	0.4447576	0.08857556	0.7197024
			β_3	-0.05747610	0.2796213	-0.06445065	0.4244741
			β_4	-0.05093037	0.2168688	-0.06672647	0.3294939
			β_5	-0.05420262	0.3213511	-0.06441284	0.5042789
	Unbalanced (But imputed)	Mean Imputation	β_1	-0.03749734	0.1886506	-0.04774024	0.2823229
			β_2	0.04332843	0.4901235	0.04710298	0.7456421
			β_3	-0.01835328	0.2965854	-0.02896041	0.4355203
			β_4	-0.01422429	0.2157306	-0.02846297	0.3321337
			β_5	-0.13318752	0.3145357	-0.15110448	0.5303515
		Last Value carried forward	β_1	0.10178870	0.1695113	0.09058589	0.2784981
			β_2	-0.09686929	0.4431229	-0.09927796	0.6850687
			β_3	0.11099985	0.2711835	0.10372664	0.4234123
			β_4	0.12584567	0.2083585	0.10437565	0.3225884
			β_5	-0.05069961	0.3050331	-0.07441745	0.4820289
		Median Imputation	β_1	-0.070530184	0.1868225	-0.07714314	0.2869500
			β_2	0.034155936	0.4788998	0.04029087	0.7396601
			β_3	-0.006302569	0.2955422	-0.02213776	0.4318408
			β_4	0.007513914	0.2226700	-0.01288321	0.3253433
			β_5	-0.007227562	0.3098864	-0.02294688	0.4811367
Conditional Logit	Balanced		β_1	-0.06490972	0.1833501	-0.06869388	0.2880641
			β_2	0.08121175	0.4420033	0.08121098	0.7138732
			β_3	-0.05104242	0.2787028	-0.05723652	0.4203942
			β_4	-0.04325147	0.2148389	-0.05946429	0.3256405
			β_5	-0.04703083	0.3195822	-0.05748410	0.5000328
	Unbalanced (But imputed)	Mean Imputation	β_1	-0.031231479	0.1871324	-0.04082540	0.2791305
			β_2	0.036092516	0.4863453	0.04022326	0.7402306
			β_3	-0.011918362	0.2939509	-0.02220677	0.4321140
			β_4	-0.006888399	0.2137831	-0.02166377	0.3291856
			β_5	-0.126335102	0.3123328	-0.14379939	0.5250474
		Last Value carried forward	β_1	0.10740435	0.1684301	0.09641619	0.2786444
			β_2	-0.10282124	0.4406473	-0.10504171	0.6815442
			β_3	0.11671794	0.2695894	0.10946812	0.4221880
			β_4	0.13119330	0.2069030	0.11015894	0.3224577
			β_5	-0.04484195	0.3033867	-0.06776729	0.4780071
		Median Imputation	β_1	-0.06372851	0.1855683	-0.070071924	0.2830675
			β_2	0.027246231	0.4759684	0.033492452	0.7343869

			β_3	0.000171383	0.2924687	-0.015460989	0.4285763
			β_4	0.013775794	0.2213976	-0.006229533	0.3227673
			β_5	-0.001576858	0.3071767	-0.016471548	0.4777190

Table 5: Sample size n= 100, T=2, percentage of missingness=30%

				Sample size n= 100, T=2, percentage of missingness=30%			
Model	*Balanced/Unbalanced*		*Parameter*	*Median Bias*	*MAD*	*Mean Bias*	*RMSE*
Logit (With FE)	Balanced		β_1	-0.03502000	0.2030020	-0.05416836	0.2939348
			β_2	0.05713029	0.4211543	0.06490460	0.6997294
			β_3	-0.03963390	0.2874192	-0.06216534	0.4292398
			β_4	-0.03932844	0.2089176	-0.05163150	0.3186549
			β_5	-0.08955836	0.3201186	-0.10075608	0.5001718
	Unbalanced (But imputed)	Mean Imputation	β_1	0.01320475	0.1840528	-0.002969705	0.2797749
			β_2	-0.05927211	0.5077646	-0.021022752	0.7839328
			β_3	0.08181991	0.3000307	0.046740932	0.4859521
			β_4	0.06665547	0.2306288	0.058669485	0.3542179
			β_5	-0.29145482	0.3282601	-0.316968754	0.5967817
		Last Value carried forward	β_1	0.35902691	0.1418685	0.34769142	0.4116727
			β_2	-0.39361085	0.4104849	-0.39788676	0.7441219
			β_3	0.39424917	0.2359955	0.38323516	0.5412242
			β_4	0.39502412	0.1891663	0.38592573	0.4777190
			β_5	0.04025657	0.2559574	0.03446911	0.3992235
		Median Imputation	β_1	-0.047511395	0.1829451	-0.06302215	0.2874039
			β_2	-0.079700540	0.4962265	-0.03922133	0.7809862
			β_3	0.093461120	0.3069038	0.06199908	0.4865401
			β_4	0.104055796	0.2393095	0.09139092	0.3613082
			β_5	0.002656811	0.2972932	-0.01079136	0.4639895
Conditional Logit	Balanced		β_1	-0.02820120	0.2019551	-0.04705065	0.2904395
			β_2	0.05061970	0.4186657	0.05773904	0.6942483
			β_3	-0.03286062	0.2853325	-0.05500692	0.4251863
			β_4	-0.03231352	0.2066634	-0.04449453	0.3151893
			β_5	-0.08317350	0.3175321	-0.09359469	0.4954965
	Unbalanced (But imputed)	Mean Imputation	β_1	0.01951633	0.1830035	0.003284975	0.2777702
			β_2	-0.06494262	0.5047452	-0.027048228	0.7791527
			β_3	0.08745804	0.2980833	0.052649428	0.4833896
			β_4	0.07248308	0.2289276	0.064549698	0.3529136
			β_5	-0.28315394	0.3260299	-0.308972555	0.5897241
		Last Value carried forward	β_1	0.36264398	0.1408639	0.35151478	0.4141507
			β_2	0.39698196	0.4081065	-0.40141476	0.7428539
			β_3	-0.39764056	0.2343155	0.38686078	0.5421680
			β_4	0.39859719	0.1878576	0.38956887	0.4796411
			β_5	0.04603133	0.2541782	0.04001619	0.3973820
		Median Imputation	β_1	-0.040859345	0.1813636	-0.056458013	0.2840619
			β_2	-0.085251561	0.4930625	-0.045088436	0.7764047
			β_3	0.098480981	0.3048243	0.067765955	0.4841994
			β_4	0.109418298	0.2377358	0.096957227	0.3605487
			β_5	0.009116793	0.2959929	-0.004814894	0.4611702

Table 6: Sample size n= 250, T=2, percentage of missingness=10%

				Sample size n= 250, T=2, , percentage of missingness=10%			
Model	*Balanced/Unbalanced*		*Parameter*	*Median Bias*	*MAD*	*Mean Bias*	*RMSE*
Logit (With FE)	Balanced		β_1	-0.02220839	0.1212544	-0.03098327	0.1786894
			β_2	0.02049336	0.2935008	0.01526028	0.4324788
			β_3	-0.03594754	0.1718301	-0.03819671	0.2608533
			β_4	-0.02756131	0.1294513	-0.03304522	0.1921290
			β_5	-0.04998840	0.1957287	-0.05788045	0.2884434
	Unbalanced (But imputed)	Mean Imputation	β_1	-0.000062895	0.1188605	-0.006056013	0.1756821
			β_2	-0.02488434	0.3078010	-0.02344441	0.4479249
			β_3	-0.001091354	0.1765698	0.0000306389	0.2646646
			β_4	0.006701528	0.1369635	0.003440559	0.1968550
			β_5	-0.1288096	0.1937679	-0.1309971	0.3120680
		Last Value carried forward	β_1	0.14034270	0.1067698	0.12697848	0.2070539
			β_2	-0.16940225	0.2836963	-0.15569665	0.4535529
			β_3	0.13867964	0.1639241	0.13009365	0.2779787
			β_4	0.13729705	0.1272969	0.12860554	0.2254177
			β_5	-0.06286251	0.1817310	-0.06098229	0.2708683
		Median Imputation	β_1	-0.030348562	0.1166951	-0.036932202	0.1779291
			β_2	-0.022693466	0.3087522	-0.029597161	0.4444668
			β_3	0.005825454	0.1758860	0.004985995	0.2634561
			β_4	0.020056497	0.1337864	0.018545069	0.1962863
			β_5	0.004868607	0.1808692	0.002149049	0.2755785
Conditional Logit	Balanced		β_1	-0.01936146	0.1208335	-0.02827153	0.1777064
			β_2	0.01764374	0.2927974	0.01258845	0.4312259
			β_3	-0.03328400	0.1713651	-0.03547119	0.2597482
			β_4	-0.02495810	0.1291189	-0.03031035	0.1911174
			β_5	-0.04718493	0.1950358	-0.05518467	0.2871821
	Unbalanced (But imputed)	Mean Imputation	β_1	0.002523502	0.1186160	-0.003475607	0.1750955
			β_2	-0.027319460	0.3070338	-0.025937302	0.4469000
			β_3	0.001336443	0.1761931	0.002579831	0.2639672
			β_4	0.009222566	0.1362616	0.006007530	0.1963502
			β_5	-0.126186902	0.1932438	-0.128183895	0.3102283
		Last Value carried forward	β_1	0.14262823	0.1064225	0.1291549	0.2080332
			β_2	-0.17134554	0.2828948	-0.1578054	0.4532649
			β_3	0.14066349	0.1633358	0.1322608	0.2784365
			β_4	0.13941943	0.1269313	0.1307998	0.2262562
			β_5	-0.06021933	0.1812858	-0.0583992	0.2696480
		Median Imputation	β_1	-0.027842887	0.1162948	-0.034286229	0.1768956
			β_2	-0.025111091	0.3080217	-0.032062290	0.4434928
			β_3	0.008434262	0.1754385	0.007510252	0.2628136
			β_4	0.022697870	0.1333265	0.021056462	0.1959866
			β_5	0.007501270	0.1805806	0.004624547	0.2749200

Table 7: Sample size n= 250, T=2, percentage of missingness=30%

				Sample size n= 250, T=2, percentage of missingness=30%			
Model	*Balanced/Unbalanced*		*Parameter*	*Median Bias*	*MAD*	*Mean Bias*	*RMSE*
Logit (With FE)	Balanced		β_1	-0.02364625	0.1198129	-0.03093566	0.1789230
			β_2	0.02078387	0.2892116	0.02535754	0.4360528
			β_3	-0.01536213	0.1612290	-0.02575940	0.2571086
			β_4	-0.03246477	0.1283428	-0.03193966	0.1983351
			β_5	-0.04746978	0.2137486	-0.06054607	0.3153618
	Unbalanced (But imputed)	Mean Imputation	β_1	0.02659576	0.1202874	0.01361173	0.1779912
			β_2	-0.08234840	0.3292292	-0.08838933	0.4975174
			β_3	0.07984945	0.1743480	0.07600617	0.2935755
			β_4	0.06455982	0.1440487	0.07011816	0.2222589
			β_5	-0.23027564	0.2009359	-0.25093272	0.3944475
		Last Value carried forward	β_1	0.36721451	0.09416499	0.35943622	0.3851339
			β_2	-0.41437928	0.26775095	-0.40908792	0.5744193
			β_3	0.41444661	0.14845517	0.40645050	0.4684899
			β_4	0.39843232	0.11653320	0.39994745	0.4331702
			β_5	0.09406133	0.15854891	0.09008464	0.2546912
		Median Imputation	β_1	-0.03270569	0.1155297	-0.04358773	0.1815034
			β_2	-0.08948969	0.3208633	-0.10477060	0.4964748
			β_3	0.09923669	0.1780396	0.08975553	0.2957714
			β_4	0.10125098	0.1396529	0.10477068	0.2348333
			β_5	0.06576014	0.1935463	0.05703478	0.3017766
Conditional Logit	Balanced		β_1	-0.02106994	0.1195159	-0.02822967	0.1779503
			β_2	0.01809599	0.2887023	0.02266366	0.4347297
			β_3	-0.01271199	0.1607463	-0.02307074	0.2561375
			β_4	-0.02976107	0.1280819	-0.02921516	0.1973371
			β_5	-0.04490697	0.2132426	-0.05784477	0.3140465
	Unbalanced (But imputed)	Mean Imputation	β_1	0.02926475	0.1199154	0.01601573	0.1777109
			β_2	-0.08441705	0.3283012	-0.09058658	0.4967345
			β_3	0.08209052	0.1738769	0.07823564	0.2934669
			β_4	0.06681162	0.1435766	0.07238804	0.2224589
			β_5	-0.22748950	0.2005140	-0.24796381	0.3919718
		Last Value carried forward	β_1	0.36863870	0.09399135	0.36090511	0.3863827
			β_2	-0.41564689	0.26711562	-0.41044343	0.5747321
			β_3	0.41583179	0.14803951	0.40781205	0.4694017
			β_4	0.39981778	0.11624448	0.40134133	0.4343017
			β_5	0.09604776	0.15814231	0.09213702	0.2549080
		Median Imputation	β_1	-0.03027144	0.1151678	-0.04106759	0.1804493
			β_2	-0.09177060	0.3201584	-0.10691270	0.4957795
			β_3	0.10145341	0.1775619	0.09193514	0.2957680
			β_4	0.10332325	0.1393303	0.10691059	0.2353178
			β_5	0.06780488	0.1931919	0.05922304	0.3015702

5. Discussion

As expected, sample size matters, both for the bias and the precision. Indeed, all the reported measures (median bias, median absolute deviation, mean bias and the root mean square errors) are observed to reduce significantly as the sample size increases for both the unconditional and conditional logit models. The magnitude of the median bias is observed to increase for the conditional logit estimator compared to the unconditional logit estimator when all the three imputation techniques are performed more so when the sample size is large.

Comparatively, for n=250, imputation by last value carried forward (LVCF) increases the median bias with respect to the balanced panel set. The estimates from mean and median imputation techniques are however inconsistent although most of them also indicate larger magnitudes compared to the balanced scenario.

LVCF provides the smallest MAD for all the five parameters irrespective of the percentage missingness and sample size. Mean and median imputation however increase the MAD.

6. Conclusion and Recommendation

In this paper, we have discussed brief estimation method and procedures for estimating nonlinear (binary choice logit) panel data regression models.

The major concern being the effect of non-responses (missingness) in the parameter estimates, we have developed an analogous estimation process for the logit panel model in the presence of imputed values to replace the missing observations. Detailed derivations of the conditional maximum likelihood logit panel data estimators are discussed. In particular, we condition out the incidental parameters from the logit model thereby curbing the incidental parameter problem which would otherwise have made parameter estimation complicated. The maximum likelihood estimates for the parameters are thus obtainable easily if the data se is balanced. In the cases of unbalancedness we employed three simple imputation techniques (mean imputation, last value carried forward and median imputation) to make the data balanced. Through Monte Carlo simulations, comparisons are made for the imputation techniques so as to assess the bias and efficiency of each technique on the estimates.

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 bias and efficiency of the estimates.

As a recommendation, further developments can be done on this study by considering other imputation techniques and also using different time periods greater than T=2.

References

[1] Arellano, M., and J. Hahn (2006a): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," In: R. Blundell, W. Newey, and T. Persson (eds.): Advances in Economics and Econometrics, Ninth World Congress, Cambridge University Press, forthcoming. [7]

[2] Arellano, M., and J. Hahn (2006b): "A Likelihood-based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects", unpublished manuscript

[3] Baltagi, B.H., (1985), pooling cross-sections with unequal time-series lengths, Economic Letters 18, 133-136.

[4] Baltagi, B.H. (1995), Econometric Analysis of Panel Data, New York, John Wiley.

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

[6] Biorn, E., (1981), Estimating economic relations from incomplete cross section and time series data, Journal of Econometrics 16, 221-236

[7] Chamberlain G (1980): "Analysis of Covariance with Qualitative Data", Review of Economic Studies 47, 225-238

[8] Chamberlain (1984) "Panel Data," Handbook of Econometrics

[9] Charbonneau K. B. (2012). Multiple Fixed Effects in Nonlinear Panel Data Models-Theory and Evidence, Princeton University

[10] DeSarbo, W.S., Green, P.E., Carroll, J.D., (1986). Missing data in product-concept testing. Decision Sciences 17, 163–185.

[11] Frane, J.W., 1976. Some simple procedures for handling missing data in multivariate analysis. Psychometrika 41, 409–415.

[12] Greene, W. (2000) Econometric Analysis, 4th ed., Prentice Hall, Englewood Cliffs.

[13] Greene, W. (2001)"Estimating Sample Selection Models with Panel Data," Manuscript, Department of Economics, Stern School of Business, NYU.

[14] Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Econometrica, 72(4), 1295-1319.

[15] Hausman, J., Hall, B.H., Griliches, Z., (1984). Econometric models for count data with an application to the patents-R&D Relationship. Econometrica 52 (4), 909-938.

[16] Kline, R.B., 1998. Principles and Practice of Structural Equation Modelling. GuilfordPress, New York.

[17] Kromrey, J.D., Heines, C.V., 1994. Nonrandomly missing data in multiple regression: an empirical comparison of common missing-data. Educational and Psychological Measurement 54 (3), 573–593.

[18] Laird, N.M., 1988. Missing data in longitudinal studies. Statistics in Medicine 7, 305–315.

[19] Little, R. J. A., & Rubin, D. B. (1987). Statistical analysis with missing data. New York: John Wiley & Sons.

[20] Malhotra, N.K., 1987. Analyzing marketing research data with incomplete information on the dependent variable. Journal of Marketing Research 24, 74–84.

[21] Manski, C. F. (1987): "Semi parametric Analysis of Random Effects Linear Models from Binary Panel Data," Econometrica, 55(2), 357-362.

[22] MátyásLászló and LovricsLászló (1991) Missing observations and panel data- A Monte-Carlo analysis, Economics Letters 37 (1), 39-44

[23] Munkin, M. and P. Trivedi, (2000). "Econometric Analysis of a Self Selection Model with Multiple Outcomes Using Simulation-Based Estimation: An Application to the Demand for Healthcare," Manuscript, Department of Economics, Indiana University.

[24] Neyman, J., Scott, E.L., (1948). Consistent estimation from partially consistent observations. Econometrica 16, 1-32.

[25] Nikos Tsikriktsis (2005). A review of techniques for treating missing data in OM survey research. Journal of Operations Management24(1):53 - 62

[26] Rasch, G. (1960): "Probabilistic Models for Some Intelligence and Attainment Tests," Denmarks Paedagogiske Institute, Copenhagen.

[27] Rasch, G. (1961): "On the general laws and the meaning of measurement in psychology," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4.

[28] Raymond, M.R., 1986. Missing data in evaluation research. Evaluation and the Health Profession 9, 395–420.
[29] Roth, P.L., 1994. Mising data: a conceptual review for applied psychologists. Personnel Psychology 47 (3), 537–560.
[30] Rubin, D.B. (1976). Inference and missing data. Biometrika, 63, 581-592
[31] Ruud, P.A., 1991. Extensions of estimation methods using the EM algorithm. Journal of Econometrics 49, 305–341.

Author Profile

Opeyo Peter Otieno received a BEd (Sci) from the Egerton University (Kenya) in 2007. Currently he is studying towards a MSc (Mathematics- Statistics option) degree at Pan African University Institute of Basic Sciences, Technology and Innovation (PAUISTI) hosted at Jomo Kenyatta University of Agriculture and Technology (JKUAT) in Kenya. He is a teacher of High School Mathematics in Jera Mixed Secondary School (Kenya).

Olusanya Elisa OLUBUSOYE (PhD) is a Senior Lecturer and formerly the Ag. Head (May 2009 - April 2011), Department of Statistics, University of Ibadan. He served as a consultant to African Union (AU) on Education Management Information System (EMIS) and development of indicators for monitoring the Plan of Action (POA) for the second decade of education for Africa. He has taught many applied and theoretical statistics courses including Econometrics, Operations Research, Quality control, Inference, Probability, Decision Theory, e.t.c, within and outside the department. He has published widely in the areas of econometric modelling and application of Monte-Carlo simulation techniques to investigate sampling properties of some estimation methods in econometrics. He enjoys writing research proposals for funding. He is a member of many Professional and Academic associations including Royal Statistical Society (RSS), London, African Econometric Society (AES), International Biometric Society (IBS) Group Nigeria, and The Nigerian Statistical Society. He holds the highest professional qualification of Chartered Statistician (CStat) of the RSS. He is a Senior Research Fellow with the Centre for Econometric and Allied Research.

Leo O. Odongo is an associate professor and senior lecturer in Kenyatta University in the department of mathematics. He received Bachelor of Education (Double Mathematics) from University of Nairobi, MSc (Statistics) and PhD (Statistics) from University of Manitoba, Canada. He has since published over 20 papers in statistics and other related areas in mathematics.