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.