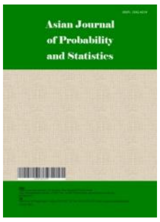

Asian Journal of Probability and Statistics

15(1): 8-14, 2021; Article no.AJPAS.71476
ISSN: 2582-0230

Incidental Parameters Problem: The Case of Gompertz Model

Ismaila A. Bolarinwa[1*] and Bushirat T. Bolarinwa[1]

Department of Statistics, The Federal Polytechnic, P.M.B. 55, Bida, Niger State, Nigeria.

Authors' contributions

This work was carried out in collaboration between both authors. Author IAB designed the study and wrote the computer program for simulation and the first draft of the manuscript. Author BTB compiled the results, managed the literature searches and wrote the conclusion. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v15i130344
Editor(s):
(1) Dr. Manuel Alberto M. Ferreira, Lisbon University, Portugal.
Reviewers:
(1) Obite, Chukwudi Paul, Federal University of Technology, Nigeria.
(2) Zhichun Yang, Chongqing Normal University, China.
Complete Peer review History: https://www.sdiarticle4.com/review-history/71476

Original Research Article

Received 10 July 2021
Accepted 11 September 2021
Published 02 October 2021

Abstract

The order of bias of the fixed effects gompertz model is studied, using Monte Carlo approach. Performance criteria are bias and root mean squared errors. For fixed N, bias is found to decrease steadily between T=5 and T=20 but exhibits a mixture of increase and decline afterwards. At each value of T involved, bias steadily decreases with increased value of N. Bias is found to be at most 123%, due to the combination of minimum of each of N and T involved. Decrease in order of bias is found to be more definite with increased N at fixed T than with increased T at fixed N.

Keywords: Bias; binary; fixed effects; inconsistency; maximum likelihood.

1 Introduction

In fixed effects panel data binary choice modeling, the individual specific effects, μ_i are assumed to be fixed and are to be estimated along with parameter vector, β . The obvious implication of such is that as the number of individuals involved in the study, N grows, the number of parameters to be estimated equally grows, with

**Corresponding author: Email: iabolarinwa@gmail.com;*

attendant loss of degrees of freedom. The fact that, maximum likelihood estimators of such binary choice models are consistent only when the number of time points, T involved in the study and not N, tends to infinity, is well documented in the literature [1-3].

Unfortunately, most times, data are available for small T and large N, since it is much easier to increase the number of individuals in a study than to increase the number of time points. The inconsistency that characterizes the estimators as N grows for a fixed T is what is described in the literature as the *incidental parameters problem* [4,5]. Unlike in the linear models, where the individual specific effects can be eliminated by some form of transformation, the situation is not same with binary choice models, which are mostly non-linear.

Within the parametric framework, a notable contribution, in form of panacea to the incidental parameters problem, based on some form of conditioning, for the logit model was made by Chamberlain [6]. This conditioning has not yielded useful results for most models. Heckman [7,8] studied order of bias for the binary probit model in static and dynamic models respectively. The order of bias for the static case is at most 10%. Results in the dynamic case indicate significant bias, which increases with the variance of the individual specific effects.

Greene [9] studied order of bias in logit and probit models and finds that at N=1000 and T=2, bias is 100%, thereby corroborating results obtained by Hsiao [2] for the logit. A few of several other works in both static and dynamic panel data models include Greene [10]; Moreira [11]; Hahn and Kuersteiner [12] and Moon, Perron and Philips [13]. More recent works include Femendez-Val and Weidner [14]; Moon and Wedner [15]; Boneva and Linton [16] and Juodis [17]. This article studies the order of bias of fixed effects gompertz model.

The remaining part of the article is organized as follows: Section 2 presents the Theoretical Framework; Section 3, presents the Methods; Section 4 presents the Results and Discussion while the last section concludes the article.

2 Theoretical Framework

The basic building block for the binary choice model of interest is the model

$$y_{it}^{*} = \beta x_{it} + \mu_i + v_{it} \qquad \text{i=1, ..., N; t=1,...T} \qquad (2.1a)$$

The observability criterion is

$$y_{it} = 1(y_{it}^{*} > 0) \qquad (2.1b)$$

μ_i is the unobserved individual specific heterogeneity;

v_{it} is the usual stochastic error term in regression;

β is a constant;

y_{it}^{*} is a latent variable observed through y_{it} .

In binary choice modeling, the primary interest is the probability that the event occurs, given the vector of regressors, x . This probability, within the context of fixed effects modeling is

$$P[y_{it} = 1/x_{it}, \mu_i] = F(\beta' x_{it} + \mu_i) \qquad (2.2)$$

The probability distribution that describes v_{it} in (2.1a) is the *tolerance*, and it determines the form of binary choice model. For the binary choice model of interest, v_{it} is standard type I extreme value (maximum). The resulting binary choice model is gompertz, defined

$$P[y_{it} = 1/x_{it}, \mu_i] = \exp[-\exp(-(\beta^{/} x_{it} + \mu_i))] \tag{2.3}$$

Designating $P[y_{it} = 1/x_{it}, \mu_i]$ by p, it follows from (2.3) that

$$-\log(-\log p) = \beta^{/} x + \mu_i \tag{2.4}$$

$-\log(-\log p)$ is the so called *link function*. Hence, the gompertz model is linear in the negative of log of the negative of the log of p.

3 Methodology

This section presents the model, data generating procedure, parameter estimation and performance criteria.

3.1 The model

The model under consideration is one described by combination of (2.1a) and (2.1b). It is a balanced 1-way fixed effects error components model with a single regressor. This model is suitable for modeling binary panel data, when the omitted individual specific heterogeneity is taken into account and inference drawn is to apply only to units involved in the study.

3.2 Data generating procedure

The latent and observed response variables are generated according to (2.1a) and (2.1b) and the exogenous variable, x_{it}, generated as obtainable in Nerlove [18]:

$$x_{it} = 0.1t + 0.5x_{it-1} + \varepsilon_{it} \tag{3.1}$$

where ε_{it} is uniformly distributed on the interval (-.5, .5). x_{i0} is chosen as $5 + 10\varepsilon_{io}$. v_{it} is generated as standard extreme value (maximum) in harmony with the tolerance requirement for gompertz model. σ_{μ}^{2} and β are each set at 1. N is set at 25, 50, 100, 150, and 200 while T is set at 5, 10, 15, 20, 25, 30, and 40. 5000 replications are performed.

3.3 Estimation of parameters

Estimation of parameters is carried out through maximum likelihood method. Under the fixed effects framework, the likelihood function for NT observations is

$$L = \prod_{i=1}^{N} \prod_{t=1}^{T} F(\beta^{/} x_{it} + \mu_i)^{y_{it}} (1 - F(\beta^{/} x_{it} + \mu_i))^{1-y_{it}} \tag{3.2}$$

So that

$$LogL = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}\log F(\beta^{/}x_{it}+\mu_i)+(1-y_{it})\log(1-F(\beta^{/}x_{it}+\mu_i))\right) \quad (3.3)$$

And

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{(y_{it}-F(\beta^{/}x_{it}+\mu_i)}{F(\beta^{/}x_{it}+\mu_i)(1-F(\beta^{/}x_{it}+\mu_i))}\right]F^{/}(\beta^{/}x_{it}+\mu_i)x_{it}=0 \quad (3.4)$$

$\frac{\partial \log L_i}{\partial \mu_i}$ can be derived by noting that L_i is the likelihood for T time points for the ithindividual.

Hence, L_i is

$$L_i = \prod_{t=1}^{T}F(\beta^{/}x_{it}+\mu_i)^{y_{it}}(1-F(\beta^{/}x_{it}+\mu_i))^{1-y_{it}} \quad (3.5)$$

$$\log L_i = \sum_{t=1}^{T}\left[y_{it}\log F(\beta^{/}x_{it}+\mu_i)+(1-y_{it})\log(1-F(\beta^{/}x_{it}+\mu_i))\right] \quad (3.6)$$

$$\frac{\partial \log L_i}{\partial \mu_i} = \sum_{t=1}^{T}\left[\frac{y_{it}}{F(\beta^{/}x_{it}+\mu_i)}.\frac{\partial F(\beta^{/}x_{it}+\mu_i)}{\partial \mu_i}+\frac{(1-y_{it})}{1-F(\beta^{/}x_{it}+\mu_i)}.\frac{\partial F(\beta^{/}x_{it}+\mu_i)}{\partial \mu_i}\right]=0$$

$$= \sum_{t=1}^{T}\left[\frac{y_{it}(1-F(\beta^{/}x_{it}+\mu_i))+y_{it}(F(\beta^{/}x_{it}+\mu_i))-F(\beta^{/}x_{it}+\mu_i)}{F(\beta^{/}x_{it}+\mu_i)(1-F(\beta^{/}x_{it}+\mu_i))}\right]\frac{\partial F(\beta^{/}x_{it}+\mu_i)}{\partial \mu_i}=0$$

$$= \sum_{t=1}^{N}\left[\frac{y_{it}-F(\beta^{/}x_{it}+\mu_i)}{F(\beta^{/}x_{it}+\mu_i)(1-F(\beta^{/}x_{it}+\mu_i))}\right]F^{/}(\beta^{/}x_{it}+\mu_i)=0 \quad (3.7)$$

where

$$F(\beta^{/}x+\mu_i)=\exp(-\exp[-(\beta^{/}x+\mu_i])) \quad (3.8a)$$

$$F^{/}(\beta^{/}x_{it}+\mu_i)=\exp(-\exp(-(\beta^{/}x_{it}+\mu_i))\exp(-(\beta^{/}x_{it}+\mu_i)) \quad (3.8b)$$

Numerical solution (using Newton-Raphson method) to (3.4) provides the parameter estimates.

3.4 Criteria for performance evaluation

The performance criteria are the bias (BIAS) and the root mean square error (RMSE), defined below:

For the experiment that is replicated r times, let us define by $\hat{\beta}_j$, the jth estimate of the true parameter value, β. Then,

$$BIAS(\hat{\beta}) = \frac{1}{r}\sum_{j=1}^{r}(\hat{\beta}_j - \beta) \tag{3.8a}$$

$$RMSE(\hat{\beta}) = \left(\frac{1}{r}\sum_{j=1}^{r}(\hat{\beta}_j - \beta)^2\right)^{\frac{1}{2}} \tag{3.8b}$$

4 Results and Discussion

Results are presented in Tables 1 to 4, attached as Appendix. For fixed N, bias decreases steadily between T=5 and T=20 but exhibits a mixture of increase and decline afterwards. This pattern is typical for all values of N. Bias is at most 123% for T=5; 39% for T=10; 31% for T=15 and 20; 34% for T=25; 32% for T=30 and 34% for T=40. Maximum bias for each T occurs at N=25 (See Table 1).

For fixed T, the situation is different as it is devoid of mixture of increase and decline; rather, a definite pattern is exhibited. At each value of T involved, bias steadily decreases with increased value of N; lower biases are associated with higher N values. For instance, at T=5, bias crashes from 123% for N=25 to 68% for N=200. This is also typical for other values of T. Bias is at most 123% for N=25; 88% for N=50; 75% for N=100; 70% for N=150 and 68% for N=200.

Median bias also decreases with decreased N at fixed T and decreased T at fixed N (See Tables 2 and 3). Median biases for various N values range between 23 and 31% while median biases for various T range between 26 and 75%. The root mean squared errors decrease at each fixed T for increasing N, same does not however, hold for fixed N and increasing T as the behaviour mimics that of biases. The mean squared errors reduce consistently between T=5 and T=15 with increasing N, this is true for each N. The reverse is the cases of T values that greater than 25 as the values increase consistently (See Table 4).

5 Conclusion

This paper investigates the order of bias of the gompertz fixed effects model. The bias is found to be at most 123% and decreases with increased N. Decrease in order of bias is found to be more definite with increased N at fixed T than with increased T at fixed N. The need to extend study to other discrete choice models is recommended.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Baltagi BH. Econometric Analysis of Panel Data. 2nd ed.Chichester: John Wiley; 2001.

[2] Hsiao C. Analysis of Panel Data. 2nd ed. Cambridge: Cambridge University Press; 2003.

[3] Greene, W.H. Econometric Analysis 5th ed.New Jersey: Prentice Hall; 2003.

[4] Neyman J. Scott, E.L. Consistent estimates based on partially consistent observations. Econometrica. 1948;16:1-32.

[5] Lancaster T. The incidental parameters problem since 1948.Journal of Econometrics. 2000;95:391-413.

[6] Chamberlain G. Analysis of covariance with qualitative data. Review of Economic Studies. 1980;47:225-238.

[7] Heckman JJ. Statistical models for discrete data, in Hsiao, C. Analysis of Panel Data 2nd ed. Cambridge: Cambridge University Press; 2003,1981$_a$.

[8] Heckman JJ. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-series data stochastic process, in Hsiao; 1981$_b$,2003.

[9] Greene WH. The behaviour of the maximum likelihood estimator of limited dependent variables models in the presence of fixed effects.2004$_a$Econometrics Journal. 2004$_a$;7:98-119.

[10] Greene WH. Fixed effects bias due to incidental parameters problem in Tobit model. Econometric Reviews. 2004$_b$;23(2):125-147.

[11] Moreira MJ. A maximum likelihood method for the incidental parameter problem. Annals of Statistics. 2009;37(6A):3660-3696.

[12] Hahn J, Kuersteiner,G. Bias reduction for dynamic non-linear panel models with fixed effects. Econometric Theory. 2011;27(6):1152-1191.

[13] Moon HR, Perron B, Philips PCB. Incidental parameters problem and dynamic panel data modeling.(Cowles Foundation Paper No. 1487); 2015.
Available:Korora.econ.yale.edu/philips/pubs/art/p1487.pdf

[14] Femendez-Val I, Weidner M. Individualand time effects in nonlinear panels with large N, T. Journal of Econometrics. 2016;192(1):291-312.

[15] Moon HR, Wedner M. Dynamic linear panel data regression models with interactive fixed effects. Econometric Theory. 2017;33:158-195.

[16] Boneva l, Linton O. A Discrete choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance. Journal of Applied Econometrics. 2017;32:1226–1243.

[17] Juodis A. Pseudo panel data models with cohort interactive effects. Journal of Business & Economic Statistics. 2018;36:77-61.

[18] Nerlove, M. Further evidence on the estimation of dynamic economic relations from a time series of cross sections.Econometrica. 1971;39:383-396.

Appendix

Table 1. Biases

T	N=25	N=50	N=100	N=150	N=200
5	1.22516	.87946	.74767	.70167	.67765
10	.39366	.34355	.31535	.30660	.30416
15	.31207	.26001	.24240	.23752	.23349
20	.30662	.25163	.23352	.22346	.22216
25	.34342	.25999	.23333	.22584	.22428
30	.32485	.26208	.24119	.23329	.23009
40	.33602	.29161	.26339	.25213	.25206

Table 2. Median biases at N

N	25	50	100	150	200
Median	.30662	0.26208	0.2424	0.23752	0.23349

Table 3. Median biases at T

T	5	10	15	20	25	30	40
Median	.74767	.31535	.24240	.23352	.23333	.24119	.26339

Table 4. Mean square errors

T	N=25	N=50	N=100	N=150	N=200
5	2.962347	1.650796	1.354399	1.271303	1.226676
10	1.026266	.905075	.849644	.832922	.824934
15	.902789	.806723	.774339	.763100	.757051
20	.902962	.795932	.763799	.750072	.745967
25	.908387	.804628	.763198	.752396	.748342
30	.920917	.804743	.771570	.758691	.753082
40	.922127	.835150	.789945	.775680	.772603

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle4.com/review-history/71476