Application of Semiparametric Non-Linear Model on Panel Data with Very Small Time Point

[1]K. Jimoh and [2]W. B. Yahya
[1]Department of Physical Sciences, Al-Hikmah University, Ilorin, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria

Abstract: *This research work investigated the behaviour of a new semiparametric non-linear (SPNL) model on a set of panel data with very small time point ($T = 1$). The SPNL model incorporates the relationship between individual independent variable and unobserved heterogeneity variable. Five different estimation techniques namely; Least Square (LS), Generalized Method of Moments (GMM), Continuously Updating (CU), Empirical Likelihood (EL) and Exponential Tilting (ET) Estimators were employed for the estimation; for the purpose of modelling the metrical response variable non-linearly on a set of independent variables. The performances of these estimators on the SPNL model were examined for different parameters in the model using the Least Square Error (LSE), Mean Absolute Error (MAE) and Median Absolute Error (MedAE) criteria at the lowest time point ($T = 1$). The results showed that the ET estimator which provided the least errors of estimation is relatively more efficient for the proposed model than any of the other estimators considered. It is therefore recommended that the ET estimator should be employed to estimate the SPNL model for panel data with very small time point.*
Keywords*: Panel data; Semiparametric Non-Linear; Model; Estimator.*

I. Introdution

Semiparametric modelling is a combination of the parametric and nonparametric approaches to construction, fitting, and validation of statistical models. There are series of non linear panel data models which are not robust to multicollinearity and heteroscedasticity according to Akeyede and Ade;leke (2015).A natural approach to modeling economic time series (panel data) with non-linear models is to define different states of the world or regimes, and to allow for the possibility that the dynamic behaviour of economic variables depends on the regime that occurs at any given point in time.Two main classes of statistical models have been proposed which formalize the idea of existence of different regimes (time, T). The popular Markov-switching models (Hamilton, 1989) assumed that changes in T are governed by the outcome of an unobserved Markov chain. Hamilton applied a 2-regime model to the US GNP growth and discovered that contractions were sharper and shorter than expansions. These models have been explored and extended in details in a number of research works (Engel and Hamilton, 1990; Hamilton and Susmel, 1994; Filardo, 1994). A different approach is to allow the regime switch to be a function of a past value of the dependent variable. Teräsvirta and Anderson (1992), Granger and Teräsvirta (1993),Teräsvirta (1994) and Akeyede and Adeleke (2015) promote a family of univariate business cycle models called smooth transition autoregressive (STAR) models.

II. Justification For The Study

The violation of the normality assumption of the error term is a bordering and challenging case to Econometricians and Statisticians. In spite of this, some researchers discovered that ignoring collinearity and unobserved heterogeneity when both are actually present in a model would lead to wrong parameter estimation (Ferdous and Bhat, 2012).Hsiao (1993, 1996) submitted that small sample size is unrealistic and in a binary choice model with a single regressor that is dummy variable and a panel in which $T(i) = 2$ for all groups, the small sample bias is 100%. For this reason, no general results exist for the small sample bias in more realistic settings. The general acceptable result is based on Heckman's (1981) Monte Carlo study of a probit model in which the bias of the slope estimator in a fixed effects model moved towards zero which was 10% when $T(i) = 8$ and $n = 100$.

On these bases, it is pertinent to develop a non-linear model that would accommodate non-normality of the error term and at the same time robust to multicollinearity and heteroscedasticity. It is also important to have a model that would be efficient when $T(i) \leq 7$. In this research work, semiparametric_ non-linear model for fitting panel data is proposed with appropriate smooth kernel function under the violation of some error structures which is robust to multicollinearity and heteroscedasticity and at the same time efficient for $1 \leq T(i) \leq \infty$.The present work therefore investigated the efficiencies of some (semi-parametric) estimators of semiparametric non-linear (SPNL) model on a set of real life data with small sample point.

III. Methodology

A semi-parametric non-linear (SPNL) model, proposed by Jimoh et al (2016), for panel data structure was developed with the incorporation of multicollinearity among the predictors and the latent variable under the violation of some assumptions on the model's error term. The error term of the model is non-normal and heteroscedastic which allow it to accommodate the positive definiteness and seasonal variation of some real life data. The model is given a

$$y_{it} = \beta_0 e^{\beta_1 \rho_{1it} X_{1it} + \beta_2 \rho_{2it} X_{2it} + \alpha_i + U_{it}}; \ i = 1, \ldots, n; \ t = 1, \ldots, T. \tag{1}$$

Hence,

$$log(y_{it}) = log\beta_0 + \beta_1 \rho_{1it} X_{1it} + \beta_2 \rho_{2it} X_{2it} + \alpha_i + U_{it} \tag{2}$$

Where;

y_{it} is the metrical response variable,

X_{1it} and X_{2it} are the predictors,

β_0 is the intercept,

U_{it} is the idiosyncratic error term,

α_i is the unobserved heterogeneity variable on U_{it},

$\rho_{1it} = \frac{cov(X_{1it}, \alpha_i)}{\sqrt{v(X_{1it}). \alpha_i}}$, is the correlation between the predictor; variable X_{1it} and the unobserved heterogeneity variable, α_i and $\tag{3}$

$\rho_{2it} = \frac{cov(X_{2it}, \alpha_i)}{\sqrt{v(X_{2it}). \alpha_i}}$, is the correlation between the predictor; variable X_{2it} and the unobserved heterogeneity variable, α_i. $\tag{4}$

IV. Method Of Data Collection

The data used in this work were initially collected during the 2006 population census by the National Population Commission (NPC), Nigeria. The dataset were collected on the percentage of penetration of telephony in Nigeria. The response variable (Y) represents the percentage penetration by households in each state, the predictors are: total number of households (X_1); number of households that have access to telephone services (X_2); and number of actual access to telephone services (X_3). All the three predictor variables were fitted non-linearly on the response variable using the semiparametric non-linear model proposed by Jimoh and Adeleke (2016). The time point of the used set of data is one (1), that is, T = 1.

V.   Results

The results were obtained using the five estimators for estimation of parameters when the SPNL model was used for fitting the set of data collected from the NPC. The results are presented and described in the following tables and graphs

Table 1: Mean Square Error (MSE) of model's parameters using different estimators (Life Data, Proposed Model)

	β0	β1	β2	β3
LSE	0.0012824	4.33E-15	1.48E-14	0.0000550
GMM	0.0012801	4.31E-15	1.48E-14	0.0000546
CUE	0.0092717	0.006698	0.0034141	0.0002497
EL	0.0080661	0.001761	0.0025709	0.0005820
ET	1.001E-14	1.001E-15	1.001E-18	1.0001E-16
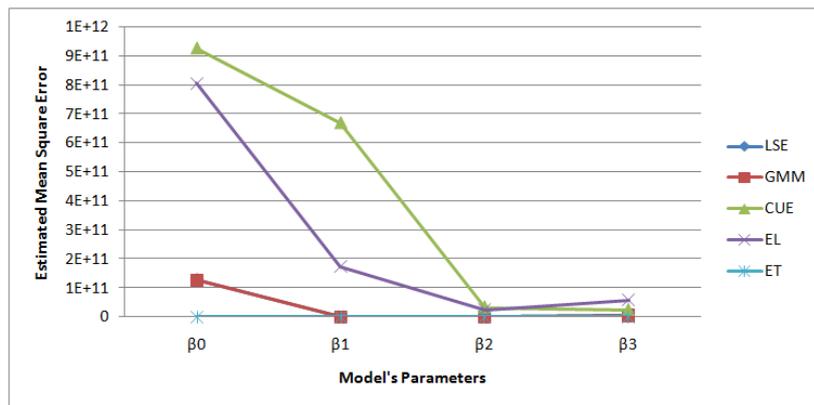
Figure 1: Line graph of Mean Square Error (MSE, in 10^{14}) of model's parameters using different estimators (Life Data,Proposed, Model)

**Table 2:** Mean Absolute Error (MAE) of model's parameters using different estimators (Life Data, Proposed Model)

	β0	β1	β2	β3
LSE	0.0011324	6.58E-10	1.22E-09	0.0002345
GMM	0.0011314	6.57E-10	1.22E-09	0.0002337
CUE	0.0030450	0.0008184	0.0018477	0.0004997
EL	0.0028401	0.0004197	0.0016034	0.0007629
ET	1.001E-17	1.001E-12	1.001E-19	1.0001E-15
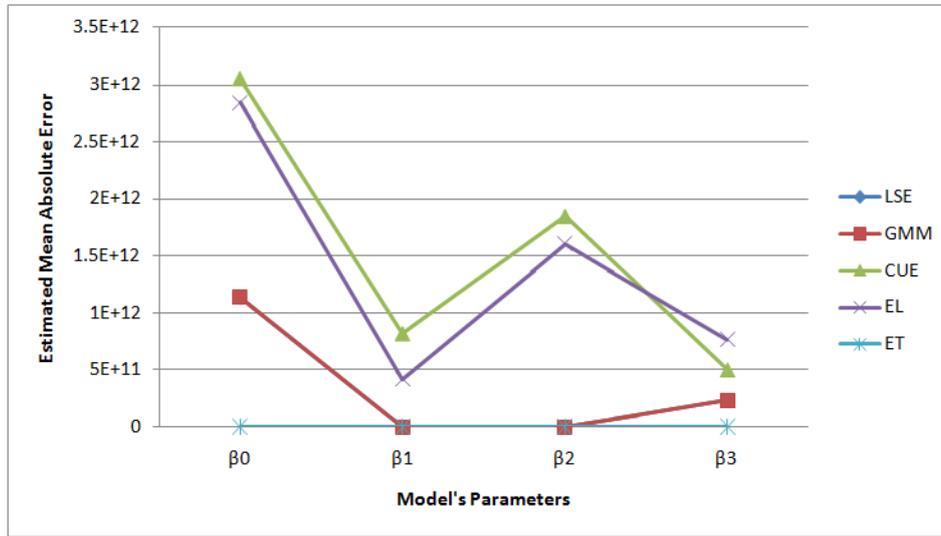
Figure 2: Line graph of Mean Absolute Error (MAE, in 10^{15}) of model's parameters using different estimators (Life Data, Proposed Model)

Table.3: Median Absolute Error (MedAE) of different parameters using different estimators (Life Data, Proposed Model)

	β0	β1	β2	β3
LSE	1.01E-19	1.001E-19	1.001E-15	1.11E-16
GMM	1.011E-19	1.11E-18	1.001E-17	1.112E-18
CUE	1.001E-19	1.011E-18	1.001E-18	1.1101E-16
EL	1.001E-19	1.001E-17	1.001E-18	1.001E-18
ET	1.001E-22	1.001E-16	1.001E-19	1.0001E-17
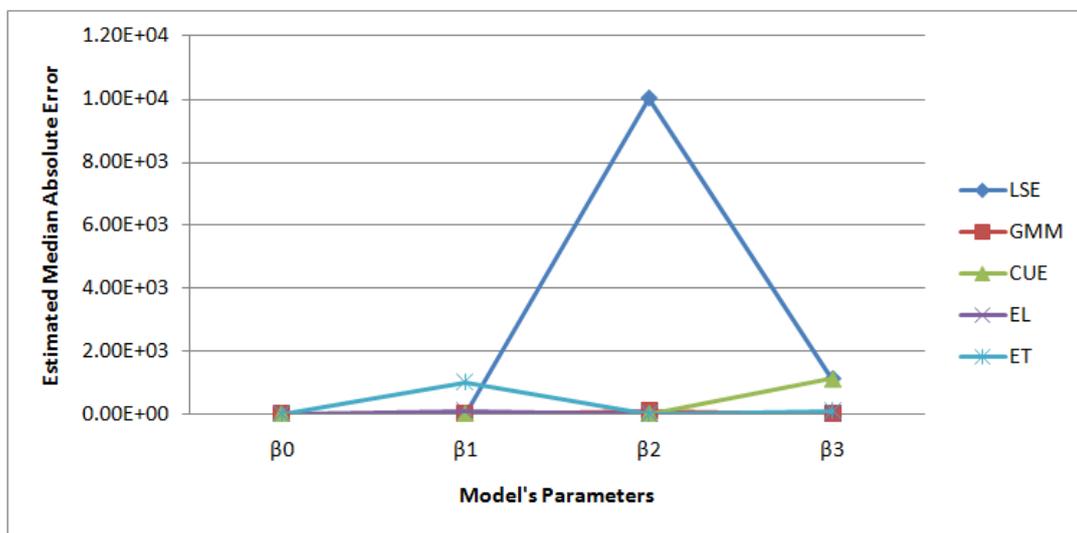
Figure 3: Line graph of Median Absolute Errors (MedAE, in 10^{19}) of the model's parameters using different estimators (Life Data, Proposed Model)

VI. Discussion Of Results Obtained From A Real Life Dataset

The efficiencies of five (semiparametric) estimators at estimating parameters of the SPNL model were investigated in this work using a real life data set. The real life data set were collected from the National Population Commission (NPC), Abuja – Census 2006. The data set were collected in order to know the rate of penetration of telephony by access to services throughout Nigeria. The response variable represents the percentage penetration by households in each state, the predictors are: total number of households; number of households that have access to telephone services; and number of actual access to telephone services.

The results from the analyses showed that ET is the best estimator when MSE, MAE, and MedAE were used as criteria for comparison among the estimators under investigation for the proposed and the existing model. The results also showed that the proposed model performed better than the existing model because the proposed model has lower MSE, MAE and MedAE at all stages of parameter estimations than the existing model.

VII.   Conclusion

Based on the various results above, it can be concluded concluded that the proposed model is efficient for modelling panel data non-linearly with or without violations of some error structures. This is because the error of estimation of each parameter using the proposed model is generally insignificant as showed in Tables 1, 2 , and 3 and as described on Figures 1, 2, and 3. As a result, it shows that the proposed semiparametric non-linear model is quite suitable to model any form of panel data, especially with small time point using any of the five estimators. Although, the results showed that the ET estimator was relatively more efficient than others, nonetheless, the use of any of the five estimators considered here for estimating SPNL model would yield good results.

References

[1].   **Akeyede and B.L. Adeleke (2015).** Nonlinear time series model. Unpublished Ph. D dissertation, University of Ilorin, Nigeria.
[2].   **Andrew J. Filardo (1994)**. Business-Cycle Phases and Their Transitional Dynamics, *Journal of Business & Economic Statistics* Vol. 12, No. 3 (Jul., 1994), pp. 299-308
[3].   **Ferdous N. and Bhat C.R, (2012)**. A multivariate ordered response model system for adults' weekday activity episode generation by activity purpose and social context. *Transportation Research Part B* 44(8-9): 922-943.
[4].   **Granger C.W.J. and T. Teräsvirta (1993).** Modelling nonlinear economic relationships. Oxford: Oxford University Press.
[5].   **Hamilton J. and R. Susmel (1994).** Autoregressive conditional heteroscedasticity and changes in regime. Journal of Econometrics 64, 307-333.
[6].   **Heckman, J. (1981)**. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process", in Manski, C. and D. McFadden, eds*., Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, pp. 114-178.
[7].   **Hsiao: C. (1993)**. *Analysis of Panel Data*, Cambridge University Press, Cambridge, pp. 159-164
[8].   **Hsiao, C (1996)**. "Logit and Probit Models," in Matyas, L. and Sevestre, P., eds., *The Econometrics of Panel Data: Handbook of Theory and Applications, Second Revised Edition.* Kluwer Academic Publishers, Dordrecht, pp. 410-447.
[9].   **K. Jimoh and B. L. Adeleke (2016)**. Robustness of Some Estimators to Multicollinearity in a Semiparametric Nonlinear Model, IOSR Journal of Mathematics (IOSR-JM) e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 6 Ver. VI PP 48-55
[10]. **Teräsvirta T. (1994).** Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of American Statistical Association* **89,** 208-218.
[11]. **Teräsvirta T. and H.M. Anderson (1992).** Characterizing nonlinearities in business cycles using smooth transition autoregressive models, Journal of Applied Econometrics **7,** S119-S136.
[12]. **Hamilton J. (1990).** Analysis of time series subject to changes in regime ,Journal of Econometrics, vol. 45, issue 1-2, pages 39-70
[13]. **Hamilton J. (1990).** A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, Econometrica, Vol. 57, No 2, Pp 357-384