IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 6 Ver. VI (Nov. - Dec.2016), PP 48-55
www.iosrjournals.org

Robustness of Some Estimators to Multicollinearity in a Semiparametric non-Linear Model

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

Abstract: *This paper proposes a semiparametric non-linear (SPNL) model that 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 estimation, for the purpose of modelling the independent variables non-linearly on metrical response variable. The performances of these estimators were examined at different levels of multicollinearity, by the use of simulated panel data. Further, the performance of the proposed non-linear model was compared with an existing non-linear model. The proposed and existing non-linear models were compared, using Absolute Error (MAE) at different sample sizes (n) and at different time points (t). For the proposed model, the performance of GMM and LS are apparently equal at time point, t = 5 and sample size, n =20 and at different combinations of n and t. For the sample size, n = 200 and time point t = 20, ET estimator was relatively better at all levels of collinearity. Overall, the ET estimator is more efficient in the proposed model compared to the existing model at all levels of multicollinearity.*

Keywords: *Panel data; Multicollinearity; Non-Linear; Model; Estimator.*

I. Introduction

Semiparametric modelling is a combination of the parametric and nonparametric approaches to construction, fitting, and validation of statistical models. In context, it is useful to review the way other approaches are used to address a generic microeconometric problem like, determination of the relationship of a dependent variable, Y to a set of conditioning variables, X given a random sample i = 1 ..... N of observations of Y and X. This would be considered as "micro"-econometric problem because the observations are mutually independent and the dimension of the conditioning variables X is finite and fixed. In a "macro"-econometric application, using time series data, the analysis must also account for possible serial dependence in the observations and a growing or infinite number of conditioning variables, e.g. past values of the dependent variable Y, which may be more difficult to accommodate. Even for microeconometric analyses of cross-sectional data, distributional heterogeneity and dependence due to clustering and stratification must often be considered; still, while the random sampling assumption may not be typical, it is a useful simplification and adaptation of statistical methods to non-random sampling.

A natural approach to modeling economic time series and 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. Roughly speaking, 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) assume that changes in T(i) are governed by the outcome of an unobserved Markov chain. Hamilton applies a 2-regime model to the US GNP growth and discovered that contractions are 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). Another estimator applied to estimation of parameters by Rui Li, Alan T. K., Wan and Jinhong You (2016) is semiparametric GMM estimator for panel data with fixed effect model. Degui Li Jin Chen and Jiti Gao (2011) also worked on time varying variables of panel data and in both cases it was discovered that the GMM estimator is more efficient than the Maximum Likelihood (ML) estimator when T(i) is small. 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 (2015) promote a family of univariate business cycle models called smooth transition autoregressive (STAR) models. These models can be viewed as a combination of the self-exciting threshold autoregressive (SETAR) and the exponential autoregressive (EAR) models. Markov-switching models imply a sharp regime switch. This assumption is too restrictive compared to the STAR models. Two interpretations of a STAR model are possible. On the one hand, the STAR model can be seen as a regime-switching model that allows for two regimes where the transition from one regime to the other is smooth. On the other hand, the STAR model can be said to allow for a continuum of

states between the two extremes (Teräsvirta, 1998). Section 2 of this paper focuses on the background of the study. Section 3 discusses the proposed model, section 4 presents the simulation scheme, section 5 also presents the results of the existing model, while discussion of results and conclusion are presented in section 6.

II. Background of the Study

This paper considered previous researches by a number of Authors. For example, Doug Walker (2008) in his Ph.D programme dissertation proposed that

$$M_i = \exp(\alpha_i + \varepsilon_i)\prod_{y=1}^{Y} X_{yi}^{\beta_y}, \quad (1)$$

where α_i is a parameter for the constant effect of brand i, X_{yi} is the value of the y^{th} variable X_y for brand i, β_y is a parameter corresponding to variable X_y, and ε_i is an error term.
Expanding the equation produces an initial brand share model,

$$m_{ijt} = \left[\frac{e^{\alpha_{ij}+\varepsilon_{ijt}} D_{ijt}^{\delta_{ij}} A_{ijt}^{\varphi_{ij}}}{\sum_{k=1}^{K} e^{\alpha_{kj}+\varepsilon_{kjt}} D_{kjt}^{\delta_{kj}} A_{kjt}^{\varphi_{kj}}} \right], \quad (1a)$$

where i = brand, j = physician, t = period, D = detailing, A = ads read in journal and α_{ij} = the constant effect of brand i with respect to physician j in a category with K brands. The parameters δ and φ represent the effects of detailing and promotional activities, respectively. The parameters δ and φ can vary by brand, physician, or both, addressing heterogeneity in physician response. He later transformed the non-linear model logarithmically to linear model as .

$$\ln(m_{ijt}) = \alpha_{ij} + \varepsilon_{ijt} + \delta_{ij}\ln(D_{ijt}) + \varphi_{ij}\ln(A_{ijt}) - \ln\left(\sum_{k=1}^{K}\exp(\alpha_{kj}+\varepsilon_{kjt})D_{kjt}^{\delta_{kj}}A_{kjt}^{\varphi_{kj}}\right) \quad (1b)$$

Another Author, Chandra R. Bhat (2000), stated that in the presence of unobserved heterogeneity, the appropriate linearized model interpretation of the PH model takes the form

$$\ln \Lambda_0(u_i) = \beta' x_i + \varepsilon_i + w_i,, \quad (2)$$

where w_i is the unobserved heterogeneity component and ε_i is the error term .

Also, Damodar N. Gujarati and Down C. Porter 2009 used an exponential model of the form

$Y_1 = \beta_1 x_i^{\beta_2} e^u$ (3)

for measure of elasticity of demand and linearized it as

$\ln Y_i = \ln\beta_1 + \ln\beta_2 x_i + u_i$, (3a)

the resulting equation was called a log-linear model.

3.0 The Proposed Model

The proposed model in this paper is stated bellow:

$y_{it} = \beta_0 e^{\beta_1\rho_{1it}X_{1it}+\beta_2\rho_{2it}X_{2it}+\alpha_i+U_{it}}$; i = 1,...,n ; t = 1,...,T; (4)

Therefore,

$log(y_{it}) = log\beta_0 + \beta_1\rho_{1it}X_{1it} + \beta_2\rho_{2it}X_{2it} + \alpha_i + U_{it}$ (4a)

$Y_{it}^* = \beta_0^* + \beta_1^* X_{1it} + \beta_2^* X_{2it} + \alpha_i + U_{it}$; i = 1, 2, ... , n ; t = 1, 2, ... , T (4b)

$\rho_{1it} = \frac{cov(X_{1it},\alpha_i)}{\sqrt{v(X_{1it}).\ \alpha_i)}}$, (4c)

$\rho_{2it} = \frac{cov(X_{2it},\alpha_i)}{\sqrt{v(X_{2it}).\ \alpha_i)}}$, (4d)

y_{it} is the response vaeiable,

$\mathbf{X_{1it}}$ and $\mathbf{X_{2it}}$ are the predictors,

β_0 is the intercept,

α_i is the unobserved heterogeneity variable and

U_{it} is the idiosyncratic error term.

Simulation Scheme

To simulate data for the study, the following schemes were designed for the generation of the panel data used for parameter estimation from the proposed model.

$Y_{it} \sim \exp(\beta)$ (5a)

$X_{1it} \sim \exp(\beta)$ (5b)

$X_{2it} \sim \exp(\beta)$ (5c)

$\alpha_i = 1$, if there exists the unobserved variable

$\alpha_i = 0$, if the unobserved variable is not present

The following values were used for the Monte Carlo Simulation
$\mu = 1, \sigma = 2, \beta_0 = 5, \beta_1 = 5, \beta_2 = 4, \beta_3 = 3$.
The Sample sizes and time points investigated are;
n = 20, n = 50, n = 100, n = 200 and n = 300; T = 5, T = 15, T = 30 with the following values of collinearities;
$\rho = 0.1$ and $\rho = 0.8$.
Parameter estimations were replicated at 1000.

III. Results of the Existing Model

A corresponding existing model (see for example, Doug Walker, 2008 and Chandra R. Bhat, 2000 and Damodar N. Gujarati and Down C. Porter, 2009) to the proposed one is of the form
$logy_{it} = log\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 \alpha_i + U_{it}$; $i=1,...,n$; $t=1,...,T$,
where, y_{it} denotes the response variable; X_{1it} and X_{2it} are the predictors; β_0 denotes the intercept; α_i denotes the unobserved heterogeneity variable; and U_{it} denotes the idiosyncratic error term. The models (Proposed and the one related to the existing models) were used for the analyses of β_0, β_1 , β_2, and β_3 for the purpose of comparison with the following results;

Table 1.2A: MAE for β_0 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

	20	50	100	200	300
LSE	1.079003	0.998006	0.899288	0.791199	0.519924
GMM	1.079113	0.998106	0.879288	0.771199	0.518924
CUE	1.078005	1.079105	1.08904	0.98143	0.908903
EL	1.07902	0.998007	0.899359	0.790904	0.500883
ET	1.079013	0.998006	0.889818	0.890796	0.500719

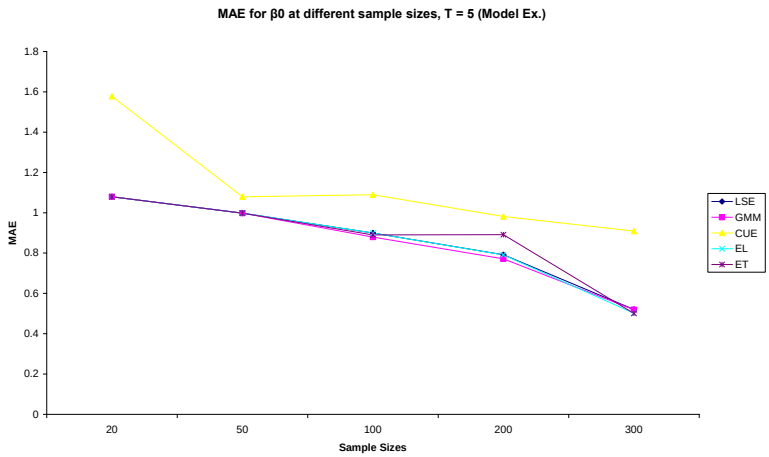

Figure 1.2A : Line graph of MAE for β_0 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

Table 1.2B: MAE for β_1 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

	20	50	100	200	300
LSE	0.166586	0.154282	0.068742	0.037618	0.017618
GMM	0.156586	0.154282	0.098742	0.039618	0.034218
CUE	0.968472	0.956773	0.971482	0.924704	0.923604
EL	0.466614	0.154322	0.068788	0.037514	0.034414
ET	0.166622	0.194303	0.038973	0.037585	0.036685

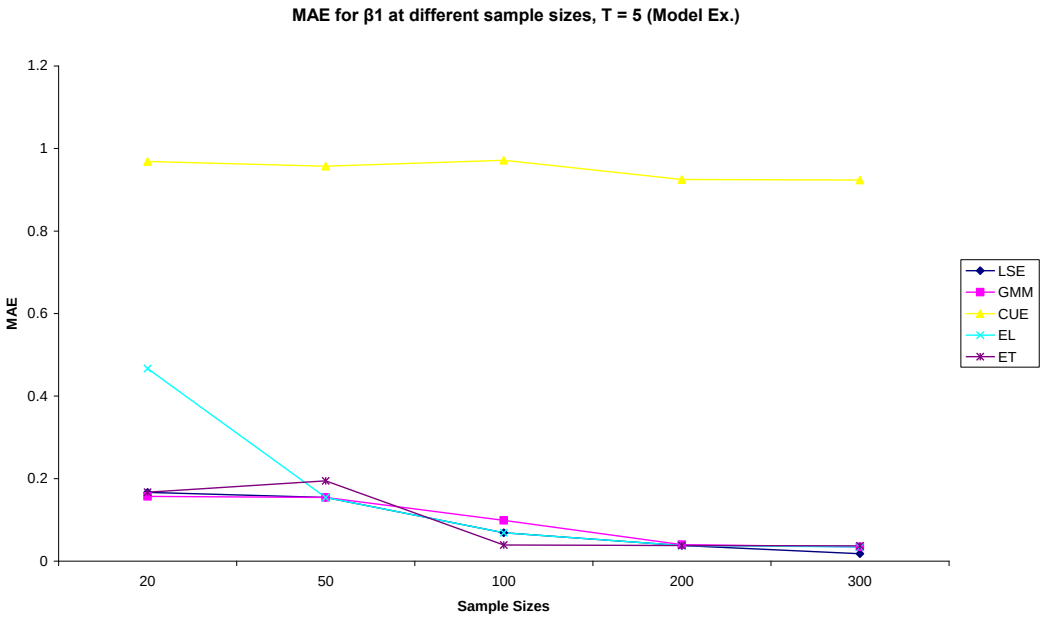

Figure 1.2B : Line graph of MAE for β_1 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

Table 1.2C: MAE for β_2 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

	20	50	100	200	300
LSE	0.137912	0.118337	0.059153	0.033365	0.037165
GMM	0.137912	0.118337	0.059153	0.033365	0.036465
CUE	0.841574	0.930439	0.862094	0.759998	0.755998
EL	0.139937	0.118336	0.059211	0.039319	0.023319
ET	0.138926	0.118287	0.059178	0.032424	0.027124

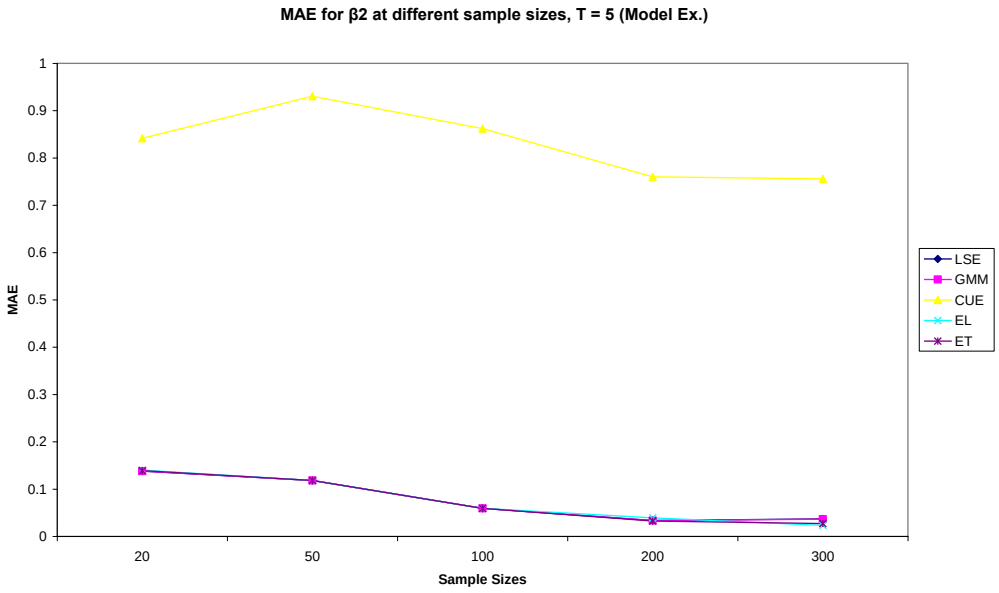

Figure 1.2C: Line graph of MAE for β_2 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

Table 1.2D: MAE for β_3 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.)

	20	50	100	200	300
LSE	0.144882	0.125438	0.062945	0.052398	0.032398
GMM	0.144882	0.125438	0.062945	0.052398	0.041998
CUE	0.563828	0.426753	0.469853	0.374001	0.272801
EL	0.14589	0.125439	0.062917	0.052382	0.041582
ET	0.145873	0.125363	0.063002	0.052392	0.041792

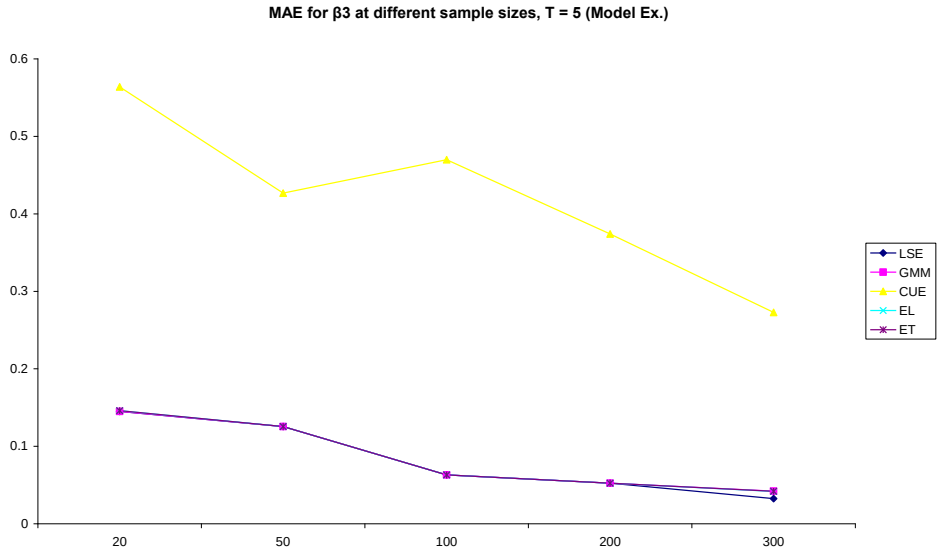

Figure 1.2D: Line graph of MAE for β_3 at different sample sizes, T = 5, $\rho = 0.1$ (Model Ex.) The corresponding results for the proposed model are presented bellow;

Table 7.2A: MAE for β_0 at different sample sizes, T = 5, $\rho = 0.1$ (Prop.Model)

	20	50	100	200	300
LSE	0.986343	0.982852	0.980412	0.978105	0.958105
GMM	0.986343	0.982852	0.980412	0.978105	0.976105
CUE	0.990042	0.999109	0.98658	0.982693	0.962693
EL	0.986437	0.982844	0.980413	0.978149	0.976149
ET	0.986449	0.982929	0.980399	0.977864	0.975264

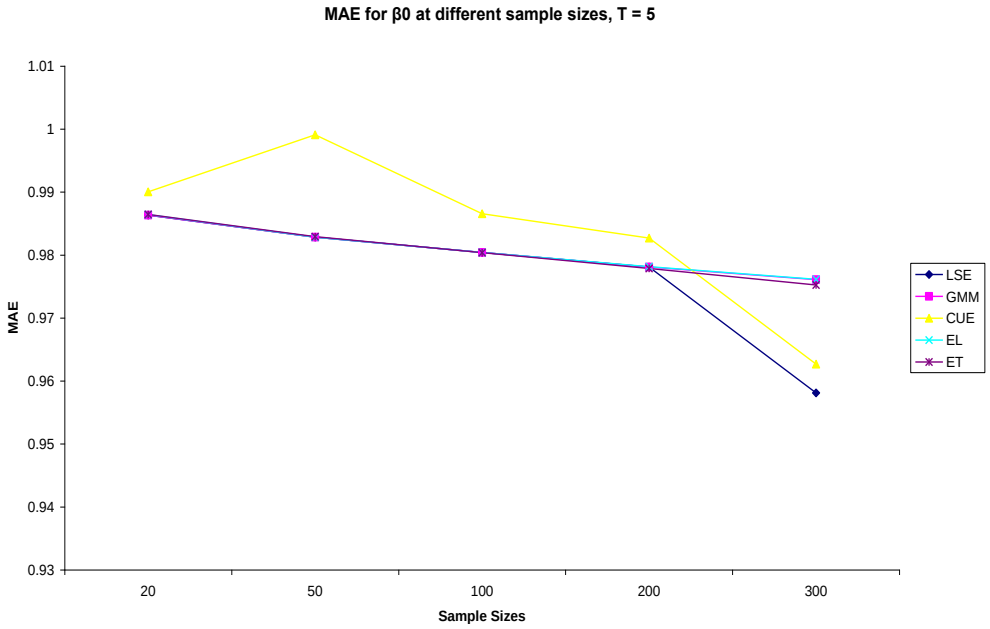

Figure 7.2A: Line graph of MAE for β_0 at different sample sizes, T =5, $\rho = 0.1$ (Prop.Model)

Table 7.2B: MAE for β_1 at different sample sizes, T = 5, $\rho = 0.1$ (Prop.Model)

	20	50	100	200	300
LSE	0.166586	0.154282	0.068742	0.037618	0.017618
GMM	0.166586	0.154282	0.068742	0.037618	0.034218
CUE	0.168472	0.156773	0.071482	0.124704	0.123604
EL	0.166614	0.154322	0.068788	0.037514	0.034414
ET	0.166622	0.154303	0.068973	0.037585	0.036685

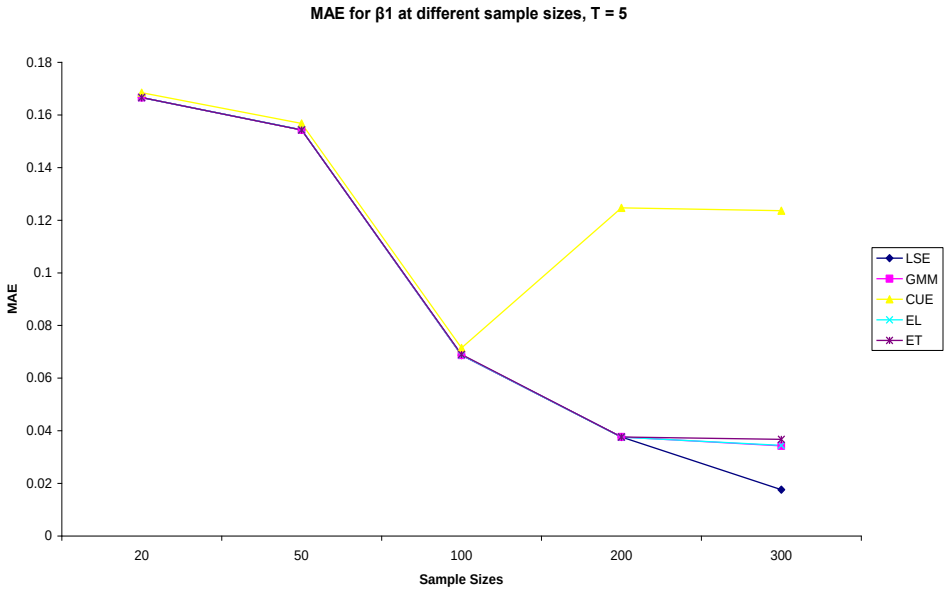

Figure 7.2B: Line graph of MAE for β_1 at different sample sizes, T =5, ρ = 0.1 (Prop.Model)

Table 7.2C: MAE for β_2 at different sample sizes, T = 5, ρ = 0.1 (Prop.Model)

	20	50	100	200	300
LSE	0.132912	0.118337	0.059153	0.039365	0.037265
GMM	0.132912	0.118337	0.059153	0.039365	0.036465
CUE	0.141574	0.130439	0.062094	0.059998	0.055998
EL	0.132937	0.118336	0.059211	0.039319	0.023319
ET	0.132926	0.118287	0.059178	0.039424	0.027424

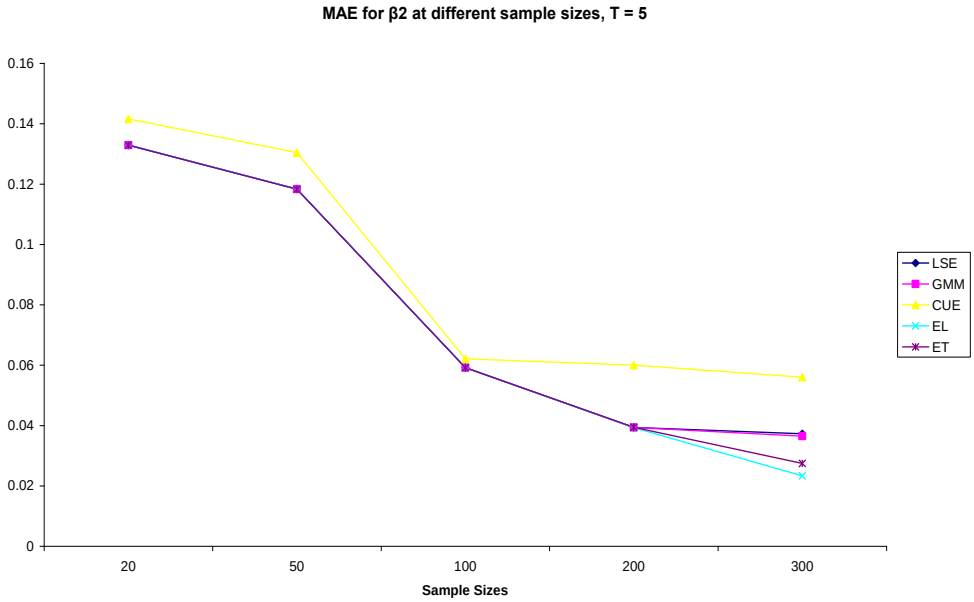

Figure 7.2C: Line graph of MAE for β_2 at different sample sizes, T =5, ρ = 0.1 (Prop.Model)

Table 7.2D: MAE for β_3 at different sample sizes, T = 5, ρ = 0.1 (Prop.Model)

	20	50	100	200	300
LSE	0.144862	0.125438	0.062945	0.042398	0.032398
GMM	0.144862	0.125438	0.062945	0.042398	0.041998
CUE	0.163828	0.126753	0.069853	0.074001	0.072801
EL	0.14489	0.125439	0.062917	0.042382	0.041582
ET	0.144873	0.125363	0.063002	0.042392	0.041792

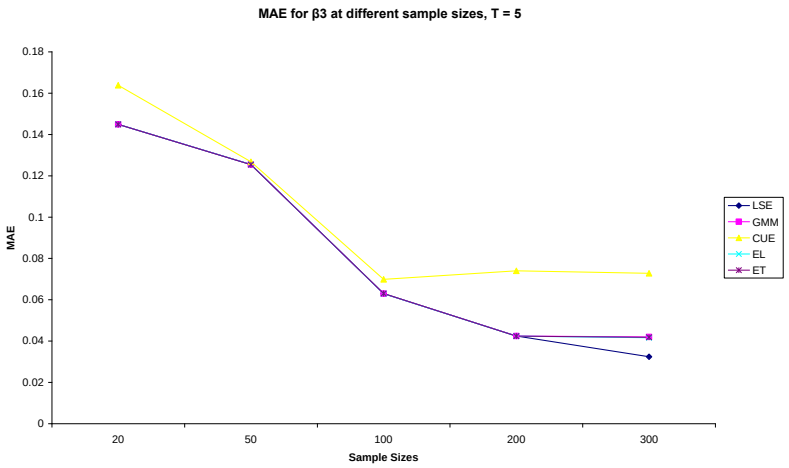

Figure 7.2D: Line graph of MAE for β_3 at different sample sizes, T =5, $\rho = 0.1$ (Prop.Model)

IV. Conclusion

We proposed a semiparametric non-linear (SPNL) model that considers the relationship between individual conditioning (predictor) variable and the unobserved heterogeneity variable before the estimation of the parameters and the main analysis. Monte Carlo Simulation was used to generate and analyse different sets of panel data with different combinations of sample sizes (n) and time point (t) together with different levels of multicollinearity using R Package. We applied different estimators like: Least Square (LS), Generalized Method of Moments (GMM), Continuously Updating (CU), Empirical Likelihood (EL) and Exponential Tilting (ET) to estimating the parameters from the simulated data.When the time point is very low, that is, T = 5 and $\rho = 0.1$ for different sample sizes, using MAE as criterion for comparison among the estimators for estimation of β_0 , CU is the best estimators when n = 20, LS and ET are the best when n = 50, GMM is the best when n = 100 and when n = 200, while ET is the best when n = 300 for the existing model. For the proposed model, LS and GMM are the best estimators when n = 20, EL is the best estimator when n = 50, ET is the best when n = 100, LS and GMM are the best when n = 200, and LS is the best when n = 300 for estimation of β_0.

When the time point is very low, that is, T = 5 and $\rho = 0.1$ for different sample sizes, using MAE as criterion for comparison among the estimators for estimation of β_1 , GMM is the best estimator when n = 20, LS and GMM are the best when n = 50, ET is the best when n = 100, EL is the best when n = 200 and LS is the best when n = 300 for the existing model. Considering the results of the proposed model, LS and GMM are the best estimators when n = 20, when n = 50 and when n = 100, EL is the best when n = 200 while LS is the best when n = 300. The best estimators for the estimation of β_2 when T = 5 and $\rho = 0.1$ are LS and GMM using MAE as criterion for comparison when n = 20, ET is the best estimator when n = 50, LS and GMM are the best estimators when n = 100, ET is the best estimator when n = 200 and EL is the best estimator when n = 300, for the existing model. LS and GMM are the best estimators when n = 20 for the proposed model, ET is the best estimator when n = 50, LS and GMM are the best estimators when n = 100, EL is the best estimator when n = 200 and when n = 300.

For the estimation of β_3 , LS and GMM are the best estimators when n = 20, T = 5 and $\rho = 0.1$, using MAE as criterion for comparison for the existing model, ET is the best estimator when n = 50, EL is the best estimator when n = 100 and when n = 200, while LS is the best estimator when n = 300. LS and GMM are the best estimators when n = 20, T = 5 and $\rho = 0.1$, using MAE as criterion for comparison for the proposed model, ET is the best estimator when n = 50, EL is the best estimator when n = 100 and when n = 200, while LS is the best estimator when n = 300. In summary, the proposed model performs better than the existing model using MAE as criterion for comparison when T = 5 and $\rho = 0.1$ for different sample sizes because MAE of the proposed model is less than the MAE of the existing model at all stages of parameter estimations.

References

[1]. **Akeyede and B.L. Adeleke (2015).** Nonlinear time series model. Unpublished Ph. D dissertation, University of Ilorin, Nigeria.
[2]. **Chandra R. Bhat (2000)**. Duration Modeling, The University of Texas, Austin Press
[3]. **Damodar N. Gujarati (2009).** Basic Econometrics, Fifth Edition. Mc Graw Hill,Singapore
[4]. **Doug Walker (2008).** "Incorporating Competitor Data into Customer RelationshipManagement". Published Ph.D. dissertation, Iowa State University.
[5]. ***Degui Li, Jia Chen And Jiti Gao*** **(2011).** Non-parametric time-varying coefficient panel data models with fixed effects. Econometrics Journal, volume 14, Pp. 387–408.
[6]. **Engel C. and J. Hamilton (1990).** Long swings in the dollar: are they in the data and do markets know it?. *American Economic Review* **80,** 689-713.

[7]. **Granger C.W.J. and T. Teräsvirta (1993).** *Modelling nonlinear economic*Relationships. Oxford: Oxford University Press.
[8]. **Hamilton J. and R. Susmel (1994).** Autoregressive conditional heteroskedasticityand changes in regime. Journal of Econometrics 64, 307-333.
[9]. **Rui Li, Alan T. K. Wan and Jinlong You (2016).** Semiparametri GMM estimationand variable estimation in dynamic panel data models with fixed effect. Journal of Computational Statistics and Data Analysis, Volume 100, Issue 6, Pp. 401-423
[10]. **Teräsvirta T. and H.M. Anderson (1992).** Characterizing nonlinearities in business cycles using smooth transition autoregressive models, Journal of Applied econometrics **7,** S119-S136.
[11]. **Teräsvirta T. (1994).** Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of American Statistical Association* **89,** 208-218.
[12]. **Teräsvirta T. (1998).** Modelling economic relationships with smooth transition regressions, in A. Ullah D.E.A. Giles (ed), *Handbook of Applied Economic* Statistics, New York, Dekker, 507- 552.