

Modeling of Per Capital Household Expenditure in Tanzania using Bayesian Two Levels Hierarchical Log-Logistic Approach: The Case Study of Dodoma Region

^[1] Milimo Mashini, ^[2] Nur Iriawan, ^[3] Heri Kuswanto.

^{[1][2][3]} *Departement of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya*
^[1] *mashinimilimo@gmail.com*, ^[2] *nur_i@statistika.its.ac.id*, ^[3] *heri_k@statistika.its.ac.id*

Abstract

In Tanzania, the fight against poverty is a long-standing agenda, various efforts and initiatives designed to eradicate poverty and increase economic growth among the citizen. However, there is evidence that the real growth over the past decade have not been reflecting in a rapid reduction in poverty rates. In this regard, the government needs analysis of household welfare or poverty. The objective of this study is to get the best model and determine the factors that explain household expenditure. Household expenditure has a hierarchical structure therefore modeled using the two-level hierarchical linear model with the characteristic of households in the first level and district characteristics at the second level. The modeling is set based on Log-logistic with three parameters (LL3) and the estimation process then accomplished by using Bayesian approach with Markov chain Monte Carlo (MCMC), and Gibbs sampling algorithms. We found that three predictors in micro model among all are statistical insignificant. These factors are age of household head, level of education and gender of head. Furthermore, in macro model all estimated parameter of the district predictors was significant at 95% credible interval. It means that the four districts predictor effected on per capita household expenditure.

Keywords: *Bayesian Hierarchical Linear Model, Log-logistic Approach, MCMC, and Per Capital HE*

1 Introduction:

One of the economic indicators that widely used to measure a sense of well-being is income per capita. However, income defined in principle as consumption plus the change in worth. In the literatures, there is common consensus that income factor applied as a measure of welfare in developed countries [1], but fortunately it appears seriously understated in less-developed countries including Tanzania [2]. Welfare issues are influenced broadly by two categories namely the behavior paradigm and the policy paradigm. The behavior paradigm related to the efforts of each household in achieving their level of welfare; within each household, there some factors that play a crucial role in contributing behavior paradigm, these factors include level of education and the number of household members. Backing to the policy paradigm category related to economic conditions, politics, and government policy. The Government policy is an external factor in the household that contributes to creating changes and improvements. Therefore, internal and external factors are influential factors of well-being in the household

Besides, the size of expenditure is more reliable as an indicator of permanent household income because spending does not fluctuate much in a short time compared to income [3]. In this regard, the Government needs analysis of per capita household income levels for the formulation, implementation, and evaluation of policies for the achievement of planned development goals.

In Tanzania, the fight against poverty is a long-standing agenda. Various efforts were taken to eradicate poverty and the increase economic growth includes the Development Vision 2025 designed in 1999, the National Poverty Eradication Strategy (NPES) designed in 1998 and Poverty Reduction Strategy Paper (PRSP) designed in 2000 [4]. All these initiatives were in place with the aim of eradicating poverty by 2025. Although the report of the Household Budget Survey shows that the incidence of poverty level declined by 6.3% from 28.2% in 2011/2012 to 26.4% in 2017/2018, still the actual causes of poverty and distribution is unknown.

Although the model would be very complex due to interaction of many parameters, representing hierarchical structure and distribution pattern of the data. Bayesian modeling approach is going to provide a better solution, flexible and easy to estimate parameters of the complex hierarchical models [5]. Therefore, it is interestingly to fit the model of per capita household expenditure and the interactions between household and districts characteristics. The present study will play a great role to create better understanding of the poverty and distribution causal in Dodoma-Tanzania.

2 Literature Review

Hierarchical levels of grouped data are commonly occurring phenomenon [6]. Many observational data collected in the biological sciences, social and economic carrier characterized with hierarchical, nested, or clustered structure [7]. Previously, it found difficult to analyses detail characteristics of these dataset due to its structural complexity. Following the development of this statistical method and its corresponding application software across many fields, it has come well appreciated. For instance, the organization of data in the education sector consist several level such as student, classroom, school, and school district levels. From this situation, the hierarchical model preferred since it used to examine the causal relationship between the measured variables at different levels in the hierarchical data structure. In this study, the hierarchical model constructed by two sub-models, micro model (level 1) and macro model (level 2).

A. Model at Level 1

In the data structure of household expenditure with two levels, let m be the number of districts and j^{th} be the number of variable within n_j as number of household within districts. The model at the first level of each district expressed as.

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \dots + \beta_{pj}x_{pnij} + \ell_j, \quad (1)$$

where

$$i = 1, 2, \dots, n_j \quad \text{and} \quad j = 1, 2, \dots, m,$$

y_{ij} = Dependent variable measured for i^{th} level-1 unit nested within the j^{th} level-2 unit,

x_{ij} = Independent variable predictors on the level-1,

β_{ij} = Regression coefficient associated with level-1 predictors,

ℓ_j = Residual vectors that assumed $\ell_j \sim N(0, \sigma^2 I)$.

The equation (1) further written in vector form such as;

$$Y_j = X_j \beta_j + \ell_j, \text{ with } y_j = [y_{1j}, y_{2j}, \dots, y_{nj}]^T,$$

$$X_j = \begin{bmatrix} 1 & x_{11j} & x_{21j} & \dots & x_{p1j} \\ 1 & x_{12j} & x_{22j} & \dots & x_{p2j} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1nj} & x_{2nj} & \dots & x_{pnj} \end{bmatrix},$$

$$\beta_j = [\beta_{0j}, \beta_{1j}, \dots, \beta_{pj}]^T, \text{ and } \ell_j = [\ell_{1j}, \ell_{2j}, \dots, \ell_{pj}]^T.$$

B. Model at level 2

In the level-2 model, the regression coefficients of the level-1 model further used as outcome variables and are related to each of the level-2 predictors. The equation model defined as follow:

$$\beta_{rj} = \gamma_{0r} + \gamma_{1r} G_{1j} + \dots + \gamma_{qr} G_{qj} + \mu_{qj} \quad (2)$$

where $r = 0, 1, 2, \dots, q$ and also expressed in vector form such that

$$\beta_{rj} = \gamma_r G_j + \mu_{qj}, \text{ with assumption } \mu_r \sim N(0, \sigma^2 I)$$

$$G_j = \begin{bmatrix} 1 & G_{11} & G_{21} & \dots & G_{i1} \\ 1 & G_{12} & G_{22} & \dots & G_{i2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & G_{1m} & G_{2m} & \dots & G_{im} \end{bmatrix}$$

$$\beta_r = [\beta_{r1}, \beta_{r2}, \dots, \beta_{rm}]^T, \text{ and } \gamma_j = [\gamma_{0r}, \gamma_{1r}, \dots, \gamma_{qr}]^T$$

$$\mu_r = [\mu_{r1}, \mu_{r2}, \dots, \mu_{rm}]^T$$

Furthermore, the equations (2) and (1) are combined to allow for the classification of variables and coefficient in terms of the level of hierarchical they affect as supported by [8]. The combined model (3) created by substitute the model (2) into model (1), and indeed defined as follow:

$$Y_j = X_j G_j \gamma + X_j \mu_j + \ell_j, \quad (3)$$

where

$$E(Y_j) = X_j G_j \gamma, \text{ is deterministic in the hierarchical model and } Var(Y_j) = X_j T X_j^T + \sigma_j^2 I_{nj},$$

$X_j \mu_j$, is a stochastic term in the hierarchical model.

C. Bayesian methods

Bayesian inference relies on Bayes theorem of probability. The Bayes model defined as

$$g(\theta | y) = \frac{f(y | \theta)P(\theta)}{h(y)}, \quad (4)$$

where θ is the parameter vector, and y is the vector of observation from sample. $P(\theta)$ represents the prior density, $g(\theta | y)$ called posterior density and $h(y)$ is defined as normalized constant. Then, the posterior distribution can be explained in the proportional form as

$$g(\theta | y) = P(\theta)L(\theta | y). \quad (5)$$

However, the posterior distribution of combined three parameters of the log-logistic Bayesian hierarchical model expressed in the form of proportional such as

$$P(\beta, \gamma, \tau_{[y]}, \tau_{[\beta]} | y) \propto f_L(y | \beta, \gamma, \tau_{[y]})P(\beta | \gamma, \tau_{[\beta]})P(\gamma, \lambda, \tau_{[y]}, \tau_{[\beta]}).$$

The estimation process of all parameters was done through repeated sampling in the form of posterior full condition distribution.

D. Distribution log-logistic three-parameter

General the Household expenditure dataset from Dodoma region identified and tested by using Easy-Fit Software. The distributions of these data have shown that the Log-logistic three-parameter is the only suitable fit for modeling the per capita expenditure data in Tanzania. The Log-logistic is a continuous positive random variable distribution with the right-skewed pattern [9]. The present study uses the LLD that defined as

$$f(y | \mu, \sigma, \lambda) = \frac{\exp\left(\frac{\ln(y - \lambda) - \mu}{\sigma}\right)}{(y - \lambda)\sigma \left[1 + \exp\left(\frac{\ln(y - \lambda) - \mu}{\sigma}\right)\right]^2},$$

where all parameters $\mu > 0$, $\sigma > 0$, $\lambda > 0$.

3 Method

A. Variables and Data Collection

Per capital expenditure model in Tanzania built by using log-logistic Bayesian hierarchical linear model. The response variable (Y) represents average expenditure per capita of HH. The level - 1 variable or predictors (X) represent the characteristics of household and the predictors (G) in the level - 2 represent characteristics of districts. The Table 1 shows the variables type considered in this study at both levels. Data collected through direct visit to the office of

Table 1 Variable description

Variable	Description	Level	Data scale
Y	Average expenditure per capita of households per month		Continuous
X ₁	Age of Household Head		Continuous
X ₂	Head household working	0 Unemployed, 1 Employed	Category
X ₃	Household farming	0 No, 1 Yes	Category
X ₄	Highest level of education of head household (HH)	0 No education, 1 Primary education, 2 Secondary and above	Category
X ₅	Number of a household member		Discrete
X ₆	Source of drinking water	0 Piped water to yard/plot, 1 Piped water into dwelling, 2 Public taps, 3 Bore hole, 4 Spring, 5 Tunker truck, 6 Surface water	Category
X ₇	The main source of cooking	0 Electricity, 1 Solar, 2 Gas, 3 Charcool, 4 Wood	Category
X ₈	Source of lighting	0 Firewood, 1 Electricity, 2 Solar, 3 Kerosine, 4 Rechargeable Lamp	Category
X ₉	Gender of Head	0 Female, 1 Male	Category
X ₁₀	Asset ownership	0 No, 1 Yes	Category
G ₁	Population density		Continuous
G ₂	The ratio of health facilities per 100,000 population		Continuous
G ₃	The ratio of education facilities per 1000 school-age population		Continuous
G ₄	The ratio of healthy person per 100,000 population		Continuous

National Bureau of Statistics of Tanzania. Both micro and macro level of dataset were obtained from the report of Household budget survey of 2017/2018. The total sample covered of household was 233 in 7 districts of the Dodoma region in Tanzania. The household expenditure (HHE) data has a hierarchical structure and the use of the hierarchical model would use to analyze different levels simultaneously in one statistical analysis. In addition, the models take account of the variance at each level of responses [10].

4 Finding And Discussion

A. Household expenditure per capita in Dodoma region

The differences of expenditure per capita between districts in Dodoma region described in Table 2. It shows that there are larger variations in the level of per capita expenditure among studied districts. Dodoma district has the highest average per capita expenditure compared to other districts with quite large differences. In addition, the coefficient of variation shows a low value.

Table 2 Descriptive Statistics of HHE per Capita of Dodoma region in 2018

Districts	Mean	Coefficient of Variation
Kondoa	82,251	56.32
Mpwapwa	144,566	132.94
Kongwa	98,414	81.09
Chamwino	70,855	69.31
Dodoma	170,794	79.76
Bahi	121,851	129.94
Chemba	62,587	38.66

B. Model distribution pattern in Dodoma region

The coefficient value of the different variations between districts can reflect the gap in welfare between residents from the point of view of expenditure/income. We found that the expenditure per capita of the population in Dodoma municipality evenly distributed among the population. The results indicating that the level of welfare of population among the districts evenly distributed during the period of study. Meaning that the variation between the higher and lower income people is not significant. We suggest that more analysis would be convenient to verify whether variation exist is significant between lower and higher income population

The difference in welfare among population in Dodoma region clearly observed from the pattern of distribution of per capita expenditure data in each district as shown by Figure 1,

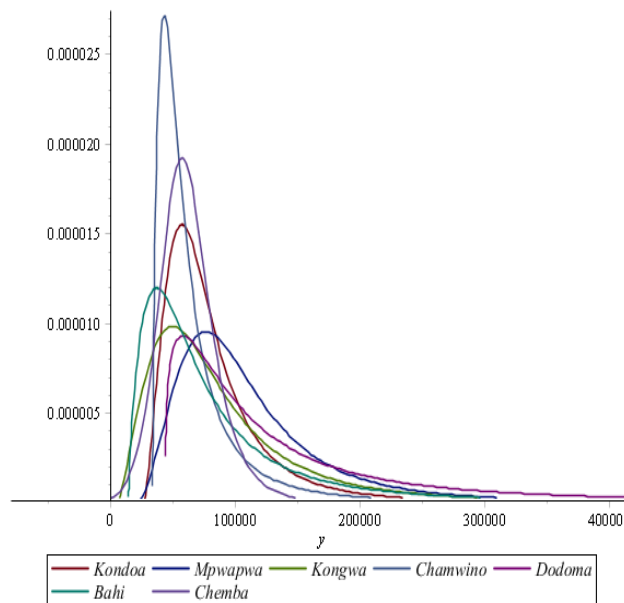


Figure 1 Distribution of HHE per capita in Dodoma region

if the parameter (shape parameter) is bigger, then the population with the lowest expenditure decreases, which means that expenditure in the area, tends to be more evenly distributed. While parameter (threshold parameter) is, describe the expenditure gap between regions

The distribution of household expenditure per capita in Dodoma region shows the expenditure gap in Mpwapwa and Chamwino districts very high compared to the other districts with the skewness of 4.6 and 3.8 respectively. Furthermore, the shift in the form of distribution in Dodoma and Kongwa shows more evenly distributed per capita expenditure compared to other districts. These districts have almost similar expenditure characteristics influenced by household characteristics and district characteristics

C. Two level Bayesian hierarchical model on per capita expenditure

Based on the three parameters log logistic with ten micro predictors and four macro predictor the estimation of the parameters of the two-level hierarchy model is carried out simultaneously using WinBUGS with 50000 iterations 50 iterations thin and additional discarded 100 burn in. The posterior estimate produces seven micro regression models that illustrate the effect of household characteristics on per capita expenditure for each household. As one of illustration significant model in micro level and combined cross-section model of level – 2 for Kongwa district as written in the equation (9) and (10).

The detail estimation results are given in annex 1 and 2. The parameters significance test in the two-level hierarchy model was done by using credible intervals, If the credible interval contains zero, then it is concluded that the parameter estimate is not significant

$$y_3 = 58.31 + 0.0085x_1 + 0.2554x_2 - 0.5188x_3 + \dots - 0.3559x_{25} + 0.1126x_{26} + 0.6849x_{27} \quad (9)$$

Interaction model for the $\hat{\beta}_0$

$$\hat{\beta}_0 = 88 + 89.77\lambda_1 + 89.99\lambda_2 + 87.9\lambda_3 + 89.99\lambda_4 \quad (10)$$

Based on the results in model (9) and (10), we found that twenty predictors among all are statistical insignificant. These factors are Age of household head, level of education and gender of head. This result obtained indeed contrast with those obtained from [11]. However, the regression coefficients, which were significant in Kongwa districts, were seven. These regression coefficients are associated with Household head working, household member, Gas as energy source of cooking, Charcoal for cooking, Wood as sources of cooking, Electricity as source of light and Ox plough as asset owned respectively. These predictors are significantly influence the per capita expenditure of household in Tanzania at Kongwa district.

Besides, the result shown that all estimated parameter of the district predictors was significant at 95% credible interval. The same results were obtained under log logistic model from the study of [12]. It means that the four districts predictors affect on per capita household expenditure in Dodoma region. Nevertheless, the estimated value of the models seems very close to each other. This results indeed are supported by the previous study of [10]. The small variation of the significant macro coefficient suggests the difference in per capital household expenditure is due to household characteristics and not districts characters. Therefore, among the macro predictors there is no variable, which has highly dominant influence on the household per capital expenditure compared to others in Dodoma region. Those four explanatory variables are population density, ratio of health facility per 100,000 populations, ratio education facility per 1000 school age-population and ration of health person per 100,000 populations.

5 Recommendation

One of the limitations in the present study is focusing only two-level hierarchical log logistic distribution. In further research, the suggestion provided to improve modeling of this field by adding levels in the hierarchical model to see the effect of three levels in improving the results in order to establish rational policies based on regional characteristics.

Several factors that are essential to influence household per expenditure are excluded due to lack of data availability and time constraints such as dependence ratio, road accessibility, and Gross domestic product per districts. Therefore, in the future study more variables considered to prevailing the detail reputation and behavior of Households' welfare in Dodoma region.

The use of the Bayes method involves decide on what prior and prior value to use therefore researchers should pay attention when they design to employs Bayes methods in modelling per capital expenditure due to its complexity in nature of using different prior distribution and prior value. In line with that, researchers should have enough knowledge in the usage of different package of software since the completion of reasonable and sound analysis incorporate more than two applications to complete analysis and obtain reflect results.

For our best knowledge, our model is strongly statistically significant. Therefore, we are suggesting to the authority especially local government to use the model as a tool for their decision making to improve welfare policies in Dodoma region for the sake of the residencies.

References

1. J. Haughton, and S. Khandker, "Handbook on Poverty and Inequality". Washington, DC 20433: The World Bank, vol. 1, March 2009.
2. T. J. Sekhampu, and F. Niyimbanira, "Analysis of the Factors Influencing Household Expenditure in South Africa Township", vol. 12, pp. 3, March 2013.
3. A. B Atkinson, and M. A. Lugo, "Growth, Poverty and Distribution in Tanzania". International Growth Centre, Oxford University, November 2010.
4. Ministry of Finance Tanzania, "The Tanzania Five Year Development Plan (FYDP II) 2020-2025". June 2019.
5. A. T. Snijders, and J. R. Bosker, "An Introduction to basic and advance multilevel modeling". London: SAGE Publication limited, 1999.
6. H. Woltman, A. Feldstain, and C. MacKay, "An Introduction to Hierarchical Linear Modeling". Tutorials in Quantitative Methods for Psychology, vol 8(1) p 52-69, January 2012.
7. H. Goldstein, "Multilevel statistical models". University of Bristol, Uk. Fourth Edition, Wily Publisher, 2011.
8. J. Gill, "Hierarchical Linear Models". Encyclopedia Social Measurements. Elsevier, 2002.
9. P. Ismartini, and N. Iriawan, "Model Criticism for Log-Normal Hierarchical Bayesian. Model Criticism for Log-Normal Hierarchical Bayesian" Published IEEE, 2013.
10. J. J. Hox, "Applied Multilevel Analysis", TT-Publikaties, Amsterdam, 1995.
11. I. Wirawati, N. Iriawan and Irhamah "Bayesian Hierarchical Random Intercept Model Based on Three Parameter Gamma Distribution". IOP Publishing, 2017.
12. B. Ngwarati, N. Iriawan, and H. Kuswanto, "Modelling of the Advanced Level National Examination Average Pass Rate in Zimbabwe using Bayesian Hierarchical Log-logistic and Normal Mixture Approach". IOP Publishing, January 2019.

Annex - 1 Regression Model for Micro-Regression Coefficients

Coefficient	Kondoa	Mpwapwa	Kongwa	Chamwino	Dodoma	Bahi	Chemba
β_0	-103.70*	37.7500*	58.310*	-37.9600*	11.930*	12.510*	-31.9600*
β_1	-0.0076	0.0035	0.0085	0.0003	0.0060	0.0153	-0.0056
$\beta_{2,1}$	-0.0443	0.0349	0.2554*	-0.0724	0.1189	0.1570	0.0434
$\beta_{3,1}$	-0.0769	-0.6380	-0.5188	0.4158	-0.1200	-0.8124	50.480*
$\beta_{4,1}$	0.1202	0.0925	0.0528	0.0643	0.0914	-0.2010	0.0168
$\beta_{4,2}$	-0.0689	-0.1181*	-0.109*	-0.1067*	-0.141*	-0.209*	-0.1487*
$\beta_{4,3}$	48840	27100	0.0000	23790	0.0000	22180	17130
β_5	113.700*	-33.770*	0.3080	19.4700	-0.4259	0.5396	-33.4200
$\beta_{6,1}$	113.900*	-33.570*	0.6786	19.3800	-0.2020	0.1474	-33.9200
$\beta_{6,2}$	48700.00	-33.02*	-0.07	23790.00	383.10*	1.62	-33.48
$\beta_{6,3}$	48920.00	27160.00	0.62	-264.30*	0.20	-0.01	-33.02
$\beta_{6,4}$	113.900*	-33.190*	1.0380	19.600	-0.2577	1.7880	-32.9900
$\beta_{6,5}$	1.55	27100.00	29980.0	28.86*	-0.24	22330.0	-99.75
$\beta_{6,6}$	-621.60*	-23.07*	-47.53*	23690.00	0.00	-2.70	16970.00
$\beta_{7,1}$	1.460	9.043	-47.22*	23800.000	-0.071	-1.923	-99.420*
$\beta_{7,2}$	1.4890	9.6570	-47.83*	29.8600*	-0.2519	-1.2020	-99.800*
$\beta_{7,3}$	0.1682	-2.0770*	0.4297	-0.6213	-0.0481	0.0900	126.400*
$\beta_{7,4}$	0.7266	-1.9530*	0.9460*	283.1000*	-0.2982	-0.3306	126.0000*
$\beta_{8,1}$	48820.00	-1.860	0.308	-0.581	-383.7*	-1.363	126.600*
$\beta_{8,2}$	0.1644	-1.949*	0.2835	-0.6231	-0.2705	-0.1233	126.40*
$\beta_{8,3}$	-0.2591	0.1251	-0.0626	0.0577	0.2843	0.3117	0.2445
$\beta_{8,4}$	623.900*	26850.000	0.632	23770.000	-0.070	1.955*	17280.00
$\beta_{9,1}$	-0.1922	0.0755	-0.1792	0.0731	0.0263	0.0666	-0.0941
$\beta_{10,1}$	48880	-0.6432	0.4356	0.6438	0.1548	1.5860*	16960
$\beta_{11,1}$	48990	27020	0.3559	23460	0.4016	-2.861*	17000
$\beta_{12,1}$	0.2207	0.1698	0.1126	0.2172	-0.1536	-0.2464	0.2173
$\beta_{13,1}$	0.1035	0.2603	0.6849*	0.4578*	-0.1290	0.1439	-0.1166

Note:* The parameter coefficients values obtained at $\alpha = 5\%$ level of statistically significant.

β_i Indicate estimate ($\hat{\beta}_i$) of the Household characters where $i = 1, 2, 3, \dots, 27$

Annex - 2 Regression Model for Macro-Regression Coefficients

Coefficient	λ_1	λ_2	λ_3	λ_4
β_0	89.97	89.99	89.94	89.96
β_1	89.99	90	89.93	89.94
$\beta_{2,1}$	89.98	89.99	89.96	89.96
$\beta_{3,1}$	89.98	89.99	89.95	89.97
$\beta_{4,1}$	89.99	89.98	89.95	89.96
$\beta_{4,2}$	90	90	89.98	89.98
$\beta_{4,3}$	89.96	89.98	89.95	89.97
β_5	89.99	89.99	89.95	89.96
$\beta_{6,1}$	89.98	89.99	89.97	89.97
$\beta_{6,2}$	89.99	90	89.97	89.96
$\beta_{6,3}$	89.98	89.99	89.94	89.97
$\beta_{6,4}$	89.99	89.99	89.94	89.97
$\beta_{6,5}$	89.97	89.99	89.94	89.96
$\beta_{6,6}$	89.98	90	89.95	89.96
$\beta_{7,1}$	89.98	90	89.94	89.95
$\beta_{7,2}$	89.98	90.01	89.97	89.96
$\beta_{7,3}$	89.98	89.99	89.95	89.97
$\beta_{7,4}$	89.99	90	89.97	89.96
$\beta_{8,1}$	89.98	89.99	89.94	89.97
$\beta_{8,2}$	89.99	89.99	89.94	89.97
$\beta_{8,3}$	89.97	89.99	89.94	89.96
$\beta_{8,4}$	89.99	90	89.97	89.96
$\beta_{9,1}$	89.98	89.99	89.94	89.97
$\beta_{10,1}$	89.99	89.99	89.94	89.97
$\beta_{11,1}$	89.99	90	89.97	89.96
$\beta_{12,1}$	89.98	90	89.94	89.95
$\beta_{13,1}$	89.98	90.01	89.97	89.96

[1] **Note:** All parameter coefficients values obtained at $\alpha = 5\%$ level of statistically significant

[2] λ_i indicate estimate of the Districts characters where $i = 1, 2, 3,$ and 4