

---

# Modeling Poverty Indices Among Crop Farmers Using Beta and Dirichlet Regression Models; A Case of Uasin Gishu County, Kenya

Jackline Chepkorir Lang'at<sup>1, \*</sup>, Thomas Mageto<sup>1</sup>, Irene Irungu<sup>2</sup>

<sup>1</sup>Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

<sup>2</sup>Department of Mathematics, Statistics and Actuarial Science, Karatina University, Kenya

## Email address:

jackiechep1992@gmail.com (J. C. Lang'at), tttmageto@gmail.com (T. Mageto), irungu.irenekareigi@gmail.com (I. Irungu)

\*Corresponding author

## To cite this article:

Jackline Chepkorir Lang'at, Thomas Mageto, Irene Irungu. Modeling Poverty Indices Among Crop Farmers Using Beta and Dirichlet Regression Models; A Case of Uasin Gishu County, Kenya. *International Journal of Data Science and Analysis*.

Vol. 7, No. 1, 2021, pp. 8-12. doi: 10.11648/j.ijdsa.20210701.13

**Received:** February 3, 2021; **Accepted:** February 10, 2021; **Published:** February 27, 2021

---

**Abstract:** Poverty and its alleviation schemes remain to be of much concern to many countries in the world. In the Sub-Saharan Africa, 41% of the population live below the extreme poverty line and in Kenya, almost 80% of the population are deemed poor. The Kenyan rural sector has a contribution of 40% to this poverty levels despite agriculture being the backbone and the main source of livelihood in the rural areas. It is in this regard that the study evaluates the household characteristics effect on Poverty indices among Crop Farmer Households. The Beta and Dirichlet regression models were used in the analysis in which the Beta regression model gave a better fit to the poverty indices data. The standardized residuals, probability plots, Chi-square test of association and the Breusch Pagan test for heteroscedasticity were used as goodness of fit evaluation tests in which levels of deprivation had a significant effect on the poverty indices among the crop farmers. Data used in the study was secondary data obtained from the Kenya National Bureau of Statistics Survey Consumption Index in Uasin Gishu County for the period March 2018 to May 2018 in which a total of 489 households were employed in the survey.

**Keywords:** Poverty, Beta and Dirichlet Regression Models, Households, Poverty Indices

---

## 1. Introduction

Poverty mitigation measures aimed at alleviating/eradicating poverty among households has been of much concern in the international arena both for developed and developing economies [1]. Globally, the poverty problem persists to cause economic havoc to many global economies and various strategies developed to curb the menace include the Millennium Development Goals by the United Nations. However, the poverty problem is a complex subject that is not self-defining as it incorporates different aspects which include social, economic and political aspects. The social aspect of poverty is the poverty of access in which households within a given a population are unable to access the basic social amenities provided by the government of the day. The economic aspect of poverty is the lack of purchasing power by the household members to buy

and provide themselves with the necessities of life. Lastly, the political aspect of poverty includes the exclusion of the households from decision making on matters that affect them and the lack of information on the available opportunities that can help better their lives. Poverty can be as a result of poor decision making within the government coupled with high illiteracy rates among the masses, lack of proper family planning methods, high levels of corruption and the unavailability of resources that can be utilized for the common good of all. Alternatively poverty was defined in monetary terms using the poverty line estimate to be US \$1.90 per day in 2011. However, recent global poverty estimates in 2015 put this figures at 735 million representing 10% of the world population [2].

Despite this decrease in global poverty levels for the period 1990-2015, the Sub-Saharan Africa showcased a different trend. Here, the poverty levels increased from 279

million people living below the extreme poverty line to 400 million people for the same period [2]. This represented a share of 50% of the global poverty index in which 41% of the Sub-Saharan Africa population was classified as being extremely poor.

In Kenya, the situation is not different from the other Sub-Saharan Africa countries. The Kenya Integrated Household Budget Survey 2015/2016 puts 80% of the Kenyan population as being poor with the rural areas contributing a 40% average compared to the pre-urban and core-urban areas (28-29) % [3]. This is despite agriculture being the backbone of the Kenyan economy and main source of livelihood among populations in the rural areas. This raises many concerns with regard to poverty levels among the farmers in the rural areas thus forming a basis for this study in evaluating household characteristics effect on poverty indices among crop farmers in Uasin Gishu County, Kenya. Here, agriculture mainly crop farming contribute to 80% of the total rural household income and food security. At least 56% of the population households engage in crop farming. The County has a growth rate of 3.8% which is higher than the national population growth rate of 2.9%. Approximately 64% of the County's population is concentrated in the rural areas whereas 36% lives in urban areas. Poverty rates in the county are estimated at 47% compared to 43.37% at the national level. Food insecurity in the county stands at 32% of households characterized by shortage of food at household level, which is mainly prominent in the months of May to August [4].

## 2. Methodology

### 2.1. Study Area and Data Source

The area of study is Uasin Gishu County one of the 47 counties in the Republic of Kenya. It has an area of 2955 KM<sup>2</sup> 290% of which is arable and a population of 11.63 Million people according to the Kenya Population and Housing Census 2019.

Data used in the study was secondary data obtained from the Kenya National Bureau of Statistics Survey on Household Consumption Index in Uasin Gishu County for the period March 2018 to May 2018 where a total of 489 households were employed in the survey.

### 2.2. Study Variables

A critical yet vital step in the modeling of poverty among crop farmers for the study was the evaluation was the evaluation of the poverty indices among the crop farmer households. Basing on average cumulative probability of access to clean water, health facilities, food, shelter and clothing, cumulative probability of an under 5 child being malnourished per household and household literacy levels per household. This was in particular to the Kenya National Housing Survey on Household Consumption Index in Uasin Gishu County, Kenya.

### 2.3. Statistical Analysis Methods

Statistical modeling of poverty indices among crop farmers involved fitting of the Beta and the Dirichlet Distributions to the data. In the modeling of poverty indices, the Classical Beta and the Dirichlet distributions were used. The beta and gamma functions of these distributions are defined as:

$$B(p, q) = \int_0^1 y_i^{p-1} (1 - y_i)^{q-1} dy \quad (1)$$

and

$$\Gamma P = \int_0^\infty t^{p-1} e^{-t} dt \quad (2)$$

respectively. Letting  $Y_i$  to be the poverty indices among crop farmers following either the Classical or Generalized Dirichlet distribution, we shall define the respective density functions as herein.

### 2.4. Review of the Statistical Models

#### 2.4.1. Classical Beta Distribution

For the poverty indices following the Classical Beta Distribution, the probability density function was defined as;

$$f(Y_i; p, q) = \frac{y_i^{p-1} (1-y_i)^{q-1}}{B(p, q)} \quad (3)$$

For  $0 < Y_i < 1$  and  $p, q > 0$  with mean and variance respectively given as  $\frac{p}{p+q}$  and  $\frac{pq}{(p+q+1)(p+q)^2}$ .

#### 2.4.2. Classical Beta Regression

For  $Y_i$  following a Beta distribution as defined in equation as 3, then the Classical Beta Regression model is defined in terms of a link function as;

$$g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} = e_i \quad (4)$$

Where  $X_{ik}$  are the household characteristics that are deemed to have an effect on poverty indices ( $Y_i$ ) among crop farmers.  $\beta_k$  are the to be estimated coefficients of each of the household characteristic.  $e_i$  is the linear predictor for the  $i^{th}$  poverty index and  $i = 1, 2, \dots, n$  where  $n$  is the sample size.  $\mu$  is the mean of the beta distribution and  $g(\mu_i)$  is a monotonic twice differentiable link function that connects the linear predictors to the response variable and is given as;

$$g(\mu_i) = \log\left(\frac{\mu}{1-\mu}\right) \quad (5)$$

#### 2.4.3. Dirichlet Distribution

For the Generalized Dirichlet distribution, We let  $Y = (y_1, y_2 \dots y_D)$  to be  $1 \times D$  positive vector with positive parameters  $(\lambda_1, \lambda_2 \dots, \lambda_D)$  and density function given as

$$F(y) = \left( \frac{\Gamma \lambda}{\prod_{l=1}^D \Gamma(\lambda_l)} \prod_{l=1}^D y_l^{\lambda_l - 1} \right) \quad (6)$$

For  $\sum_{l=1}^D y_l = 1$  and  $\sum_{l=1}^D \lambda_l = \lambda$

The mean and variance are respectively given as

$$\frac{\lambda_i}{\lambda}$$

And

$$\frac{\lambda_i}{\lambda} \left(1 - \frac{\lambda_i}{\lambda}\right)$$

$$\lambda + 1$$

#### 2.4.4. Dirichlet Regression

For  $Y_i$  following a Dirichlet distribution as defined in equation 6, then the Dirichlet Regression model is defined in terms of a link function as;

$$g(\mu_i) = \log(\lambda_i(X_k)) = \lambda_i + \sum_{j=1}^J \beta_{jk} X_{ijk} \quad (7)$$

For  $E(y_i) = \frac{(\lambda_i(X_k))}{\sum_{m=1}^D \lambda_m(X_k)}$  where  $\beta_k$ ,  $X_{(ik)}$  and  $g(\mu_i)$  are as with the classical beta regression

#### 2.5. Model Validation

This was via the AIC & BIC model selection techniques. The crop farmer household characteristics (the place of residence of a household, household size, the number of adults in a household, household crop income, household total expenditure and household levels of deprivation) were evaluated using the Beta and Dirichlet Regression Models. [14]. Prior to modeling of the Poverty Indices data, the study model assumptions were given consideration select the model that best fits the data. If  $p$  is the number of parameters and  $n$  the sample size then the AIC/BIC are given as;

$$AIC = -2 \log(L) + 2p \quad (8)$$

$$BIC = -2 \log(L) + \log(n) p \quad (9)$$

The model with the smallest AIC or BIC values was deemed to give a better fit in modeling poverty indices among crop farmer households.

## 3. Results and Discussion

### 3.1. Exploratory Data Analysis

To aid in the data analysis, a total of 489 households used in the KNBS (Kenya National Bureau of Statistics) Survey on Household Consumption Index in Uasin Gishu County were employed in the study. This was for the period March 2018 to May 2018. The descriptive statistics and the fitted model coefficients were used to give an insight into the modeling of poverty indices among crop farmers in Uasin Gishu County.

### 3.2. Estimation of the Model Parameters

The AIC & BIC of the fitted models were used to determine the model that gave better parameter estimates for the total expenditure data. Table 1 gave a summary of the fitted Model AIC & BIC values.

Table 1. The summary of the DIC values of the models.

|                      | AIC       | BIC       |
|----------------------|-----------|-----------|
| Beta Regression      | -322.0739 | -4522.595 |
| Dirichlet Regression | -302.7231 | -4421.003 |

The Beta Distribution gave lower AIC & BIC values thus giving an indication of its capability of modeling Poverty indices (proportional data) better than the Dirichlet Distribution.

### 3.2.1. Summary Statistics

From the descriptive statistics, the minimum and maximum poverty indices were 0.0010 and 0.5330 respectively, which gave an indication that the indices lied in the interval  $[0, 1]$  and there was no catastrophic household poverty index. On whether the indices followed a Beta/Dirichlet distribution as a pre-requisite to fitting the Beta/Dirichlet Regression models, the Breusch Pagan and Skewness were used. The Breusch Pagan test gave a p-value of 0.00648 which was less than 0.05 thus giving an indication of the presence of heteroscedasticity in the data. Skewness of 1.0839 confirmed the presence of asymmetry in the data. This presence of heteroscedasticity and asymmetry in the data provided grounds for the fitting of the Beta and Dirichlet Regression Models. The histogram was used to give a graphical visualization of the data and the associated density curve and the parameter estimates of the fitted models were obtained by the use of the maximum likelihood approach. Since the Beta Regression gave a better fit than the Dirichlet Regression, it's estimated and predicted model parameters were discussed.

Here levels of deprivation (Deprivation) had a significant effect of 0.1777 unit increase for the poverty index per household for a unit increase in Deprivation. The minimum and maximum predicted poverty indices were 0.01478 and 0.67909 respectively and this was skewed to the right.

The goodness of fit of the models was via the Standardized residuals, Quantile-Quantile plots and the Pearson Chi-Square test of association for Estimated and Predicted Household Poverty Indices. The Beta residuals had a median quantile of -0.0244 which was close to zero and lower model confidence intervals compared to the those of the Dirichlet Model (-0.3994). This ascertained the prowess of the Beta Regression Model in modeling poverty indices compared to the Dirichlet Regression Model. The Pearson Chi-Square P-value of  $4.4 \times 10^{-10}$  which was close to 0 implied presence of association between the observed and expected poverty indices.

### 3.2.2. Poverty Indices Prediction

Figure 1 gave a graphical visualization of the predicted household poverty indices among the crop farmer households in which the predicted poverty indices were skewed to the right. This confirmed the presence of heteroscedasticity in predicted poverty indices as with the modeled poverty indices.

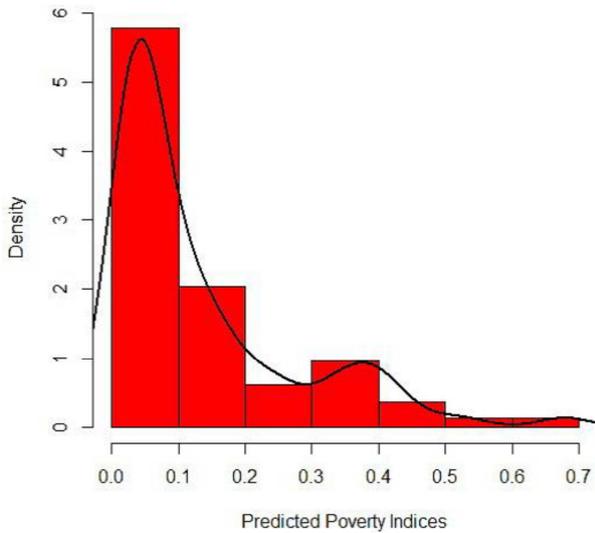


Figure 1. Histogram of predicted poverty indices.

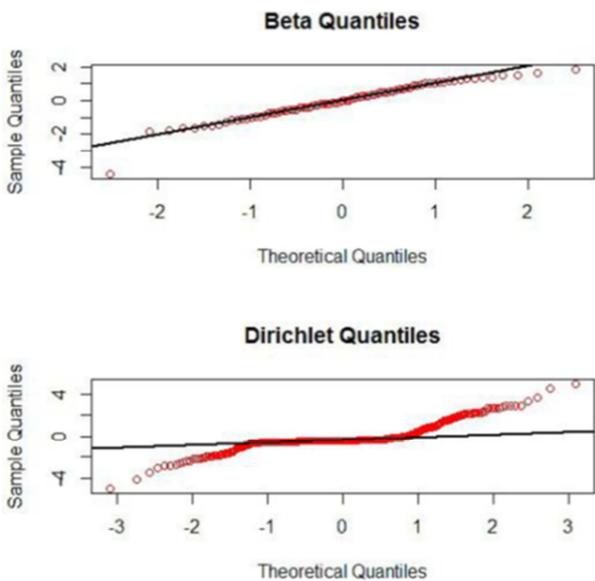


Figure 2. Fitted Model Quantiles in the distributions.

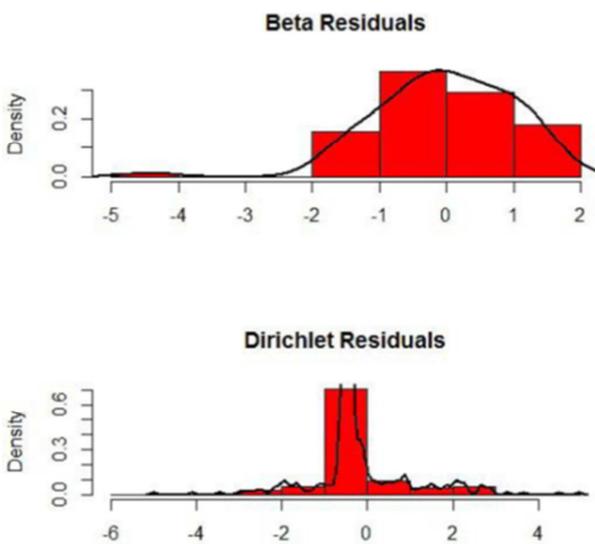


Figure 3. Histogram of residuals in the Beta and Dirichlet models.

Table 2. Dirichlet Regression Coefficients.

|                 | Estimate                 | Std. Error                | Z-value | Pr (> z )                |
|-----------------|--------------------------|---------------------------|---------|--------------------------|
| Intercept       | -3.789                   | 0.8216                    | -4.612  | 3.98 × 10 <sup>-06</sup> |
| Residence       | 0.1396                   | 0.1752                    | 0.797   | 0.4257                   |
| HH. Size        | 0.03072                  | 0.1456                    | 0.211   | 0.8329                   |
| HH. Adult       | -0.03302                 | 0.1817                    | -0.182  | 0.8557                   |
| C. Income       | 4.29 × 10 <sup>-05</sup> | 1.539 × 10 <sup>-04</sup> | 0.279   | 0.7801                   |
| HH. Expenditure | 4.54 × 10 <sup>-06</sup> | 2.835 × 10 <sup>-05</sup> | 0.160   | 0.8727                   |
| Deprivation     | 0.03693                  | 0.01694                   | 2.180   | 0.0293                   |

### 3.3. Related Work

As a basis for this research, the study acknowledges the past studies on poverty modeling and puts an emphasis on using a statistical model that captures the asymmetry and heteroscedasticity that are usually present in poverty indices data. This is as opposed to [1, 2, 5] who used the least square regression which could not test and account for the heteroscedasticity in the data. The study also acknowledges the multi-dimensional complex nature of poverty thus the need to have a response that captures all the aspects of poverty. Its not efficient to use a poverty line as it is sometimes difficult to measure. This is as with [6-8]. Its in this regard that the study used the Household Poverty Index per household aimed at capturing all the societal aspects of poverty. There is also need to dig deep into specific crop farmer household characteristic that influence poverty indices among the crop farmer households. Previous literatures provide global measures of poverty thus the need of narrowing down on the them factors that influence poverty among households. This was as with [9-12].

## 4. Conclusion and Recommendation

To help better the lives of people who live in the rural Kenya and who depend solely on subsistence farming as a source of livelihood, it is imperative for the Kenyan government to come up with policies aimed at increasing agricultural productivity which can lead to the development of an ecological, reasonable and varied agricultural sector. This will in turn lead to an improved agricultural output translating to high revenues for farmers thus helping curb the poverty menace among crop farmers. In the modeling of Household Poverty Indices as a measure of poverty levels among crop farmer households, a total of 489 households from the Kenya National Bureau of Statistics Survey on Household Agricultural Consumption Index in Uasin Gishu County, Kenya were employed in the study. The study gave the Beta Regression Model a consideration in the modeling poverty indices among crop farmer households in comparison to the Dirichlet Regression Model. This is due to reason that the Beta regression accounts better for the Heteroscedasticity in this kind of data-sets than the Dirichlet Regression Model [15]. Modeling household characteristic effect on poverty indices among crop farmer households is a crucial step in developing poverty alleviation schemes for developing countries. The various statistical approaches to

modeling of proportional data needs to be applied in the modeling of poverty indices data-sets and the associated dispersion therein. This study gave an application of the Beta and Dirichlet Regression Models in the modeling of Poverty indices among crop farmers. Other methodologies that incorporate Bayesian Inference and the Non-Parametric inference can be looked into with regard to modeling longitudinal data with an application to consumption index data-sets. The agricultural sector remains a key driver of poverty reduction in Kenya thus necessitating the need for research in this area to facilitate the formulation and implementation of poverty reduction/alleviation policies. This would include government incentives aimed at reducing consumption expenditure among crop farmers including but not limited to; provision of quality farm inputs at cheaper prices, the farmers' access to education, provision of energy, Proper housing and improved infrastructure.

---

## References

- [1] Sabina, A., Jame, E. F., Suman, S., Maria, E. S., Jose, M. R. & Paolla, B. (2015). Multidimensional Poverty Measurement and Analysis: Chapter 10- Some Regression.
- [2] Measurement and Analysis. Oxford Poverty Human Development Initiative (OPHI) Working Paper No. 91.
- [3] Sidikat, L., Gafar, T. & Usman, A. (2009). Determinants of Poverty in Sub-Saharan Africa. *African Research Review*, 3 (2); 162-177.
- [4] Haradhan, K. M. (2013). Poverty and Economic Development of Kenya. *International Journal of Information Technology and Business Management*, 18 (1); 72-82.
- [5] Afrobarometer Round 7 Survey Results in Kenya, (2020). Institute of Development Studies, University of Nairobi-Kenya. [www.afrobarometer.org](http://www.afrobarometer.org).
- [6] Nekesa, M. P. (2015). Determinants of rural poverty in Kenya: The cash crop growing. University of Nairobi Publications. <http://erepository.uonbi.ac.ke>.
- [7] Apata, T. G., Apata, O. M., Igbalajobi, O. A. & Awoniyi, S. M. O. (2010). Determinants of rural poverty in Nigeria: Evidence from small holder farmers in South-western, Nigeria. *Journal of Science and Technology Education Research*, 1 (4); 85-91.
- [8] Baiyegunhi, L. T. S. & Fraser, G. C. G. (2010). Determinants of Household Poverty Dynamics in Rural regions of the Eastern Cape Province, South Africa. 3rd African Association of Agricultural Economists (AAAE). University of KwaZulu-Natal.
- [9] Almas, H. & Seyoung, C. (2017). The Effects of Lifetime Work Experience on Incidence and Severity of Elderly Poverty in Korea. Institute of Labor Economics. [www.iza.org](http://www.iza.org)
- [10] Schneider, K. & Gugerty, K. (2011). Agricultural Productivity and Poverty Reduction: Linkages and Pathways. *The Evans School Review*, 1 (1).
- [11] Namara, E. R., Godswill, M., Fitsum, H. & Seleshi, B. A. (2017). Rural poverty and inequality in Ethiopia: does access to small-scale irrigation make a difference? International Water Management Institute. [r.namara@cgiar.org](mailto:r.namara@cgiar.org).
- [12] Chakra, P. A. & Roberto, L. G. (2012). The Impact of Remittance on Poverty and Inequality: A Micro-Simulation Study for Nepal. National Graduate Institute for Policy Studies 7-22-1 Roppongi, Minato-ku, Tokyo, Japan 106 8677. <http://www.grips.ac.jp/rcenter/wp-content/uploads/11-26.pdf>.
- [13] Sinnathurai, V. & Brezinova, O. (2012). Poverty Incidence and its Determinants in the Estate Sector of Sri Lanka. *Journal of Competitiveness*, 4 (1); 44-55.
- [14] Hijazi, R. H., & Jernigan, R. W. (2009). Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability & Statistics*, 4 (1), 77-91.
- [15] Camargo, A. P., Stern, J. M., & Lauretto, M. S. (2012, May). Estimation and model selection in Dirichlet Regression. In *AIP Conference Proceedings 31st* (Vol. 1443, No. 1, pp. 206-213). American Institute of Physics.