

# On Skew-Normal Model for Economically Active Population

Olosunde Akinlolu A  
Department of Statistics  
University of Agriculture, Abeokuta  
P.M.B. 2240, Italy.  
akinolosunde@yahoo.ca

## Abstract

The literature related to skew-symmetric distribution have grown rapidly in recent years but at the moment no publication on its applications concerning the description of economically active data with this type of probability models. In this paper, we provided an extension to this skew-normal distribution, which is also part of the family of skewed class of normal but with additional shape parameters  $\delta$ . Some properties of this distribution are presented and finally, we considered fitting it to economically active population data. The model exhibited a better behaviour when compared to normal and skew normal distributions.

**Keywords:** Skew normal model, Hidden truncation method, Economically active data, Shape parameter, Probability density function.

## 1. Introduction

In this paper, we introduced extension to the classical Azzalini's skew normal distribution; with additional shape parameter which made it suitable for modelling and somewhat generalized the model. We investigated the applicability of extended skew normal model as possible model for economically active population. This probability model was originally introduced without the idea of their use in economically active population modelling; see Azzalini (1985), Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). However, upon further reflection, it turns out that their structure may be appropriate for this purpose.

A real random variable  $X$  has a standard skew-normal distribution with a probability density function  $2\phi(x)\Phi(\lambda x)$ . This is a family of distributions on the real line which includes the normal model  $\phi(x)$ , but with shape parameter  $\lambda$  which allows flexibility in the skewness of the density depending on the value of  $\lambda$  and  $\Phi(\cdot)$  is the cumulative distribution function of the normal distribution. The development of this theoretical field has followed the general tendency in Statistics towards more flexible methods to represent features of the data, as adequately as possible, and to reduce unrealistic assumptions as the normality that underlies most methods of univariate and multivariate analysis. The key advantage of skewing a symmetric distribution  $\phi(\cdot)$  is that in doing so it is possible to retain some of the properties of  $\phi(\cdot)$ , which are often well known and further acquire new properties. The skew normal model has been found useful in fitting some set of data in both univariate and multivariate cases for instance; for modelling financial data see; Adcock (2004), Chen, et.al (2003) and De Luca, et. al. (2005), also Bolfarine and Lachos (2006) examined the regression model under the assumption of skew normal distribution.

It is worth noting that Gupta and Gupta (2004), presented a generalization to this model; both mathematical and statistical properties of this generalization were examined and also applied. In this study, we considered introducing new shape parameter,  $\delta$  to Azzalini, (1985) skew normal distribution which gave the extended type with additional shape parameter  $\delta$ . This model has an advantage of modelling data that required moderate value of  $\lambda$  (earlier considered by Azzalini, 1985), for which in the case of skew normal distribution the side with shorter tail has its side linear in shape which makes it look like half-normal distribution, this make it unsuitable for modelling economically active population data which we are considering in particular in this study. Therefore, there is need to introduce the additional shape parameter  $\delta$ , to alleviate the short coming of the Azzalini's model.

## 2. Derivation and Important Properties of the Density Used in Fitting Economically Active Population Data

The existence of the model used in fitting the economically active population data is justified by the fact that, recalling the results from Ellison (1964) and Zack (1981, pp. 53-54). They proved that if  $Y \sim N(\mu, \sigma^2)$ , then

$$E[\Phi(Y)] = \Phi\left(\frac{\mu}{\sqrt{1 + \sigma^2}}\right) \quad (1)$$

Where  $\Phi(\cdot)$  is the distribution function of the normal random variable, following the result (1), also recalling the construction of skew normal model,  $2(\cdot)\Phi(\lambda x)$ , then, we can immediately express the standardized real random variable  $X$  which has extended skew normal distribution as

$$f(x | 0, 1, \lambda, \delta) = 2\phi(x)\Phi\left(\frac{\lambda x}{\sqrt{1 + \delta x^2}}\right), \quad \lambda \in \mathbb{R}, \delta > 0 \quad (2)$$

where  $\phi(\cdot)$  is the normal density function and  $\Phi(\cdot)$  is distribution function of the normal random variable,  $\lambda$  and  $\delta$  are shape parameters. Noting that this extended skew normal distribution becomes Azzalini (1985, 1986) skew normal when  $\delta=0$ , and normal distribution when  $\lambda=\delta=0$ . Therefore, further generalizes the model.

To apply this model, we introduce the location ( $\mu$ ) and scale ( $\sigma$ ) parameters then we have (3).

$$f(x | \mu, \sigma, \lambda, \delta) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left[\frac{\lambda(x - \mu) / \sigma}{\sqrt{1 + \delta(x - \mu)^2 / \sigma^2}}\right], \quad \lambda, \mu \in \mathbb{R}, \quad \delta, \sigma > 0 \quad (3)$$

Following the procedure adopted by Azzalini (1985) we can find the moment generating function (MGF) of (3) by noting that if  $Z$  has standard extended skew normal distribution, then  $Z^2$  is  $\chi_1^2$  distributed, the implication of this is that the even moments of  $Z$  are equal to the moments of the standard normal distribution,

Just as it was obtained in the skew normal case. For the moments generating function (MGF) of  $X$  distributed as (3) we have;

$$M_x(t) = E[\exp(tX)] = 2 \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\} E\left[\Phi\left(\frac{\lambda(Y - \mu + t)}{\sqrt{\sigma^2 + \delta(Y - \mu + t)^2}}\right)\right] \quad (4)$$

Where  $Y \sim N(\mu, \sigma)$ .

To obtain the first, second, third and fourth moments (4) is differentiated as applicable to each moment, not shown because there is no implicit expression for them. The cumulative distribution function (cdf) is not also in close form. The extension to Azzalini (1985) is considered because of the deficiency in the model which make it to be inadequate for the data in consideration, for a moderate value of  $\lambda$ , the skew normal model has its mass concentrated on either the negative or positive side as determined by the sign on  $\lambda$ , this makes the distribution appear like half normal. For the case of economically active data, the extension to the skew normal behaves better on the side with shorter tail.

### 3. Economically Active Population Data.

#### 3.1 Definitions and Classifications

An **economically active population** are subset of the population who are economically vibrant to engage in wide range of activities, ranging from education, employment, social activities /vices, politics, crime, war e.t.c., they are the major participants in the economic growth of any nation. The set include in the analysis are;

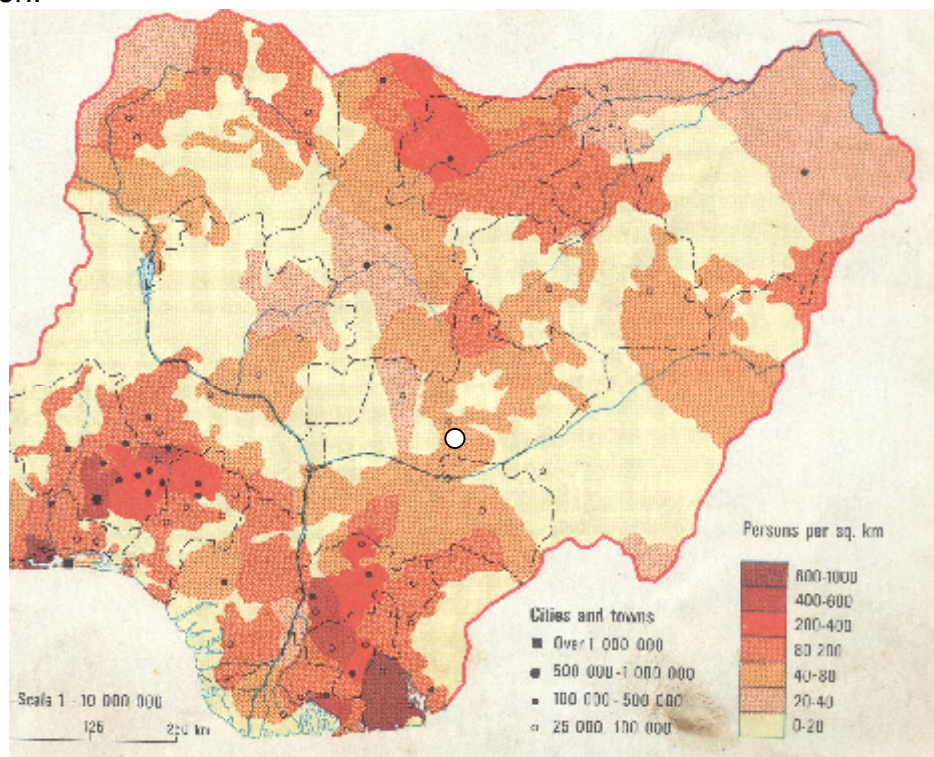
**Labour force** – persons (aged 10 years and older and persons whose age is unknown) who were considered employed or unemployed or who were conscripts in the week preceding the census. Some set of people we termed economically inactive population were excluded in the survey– persons aged 10 years and older who did not want to or could not work for some reason but who were not considered as unemployed or who were not conscripts in the week preceding the census. Under this we have;

1. **Student (pupil)** - person who was acquiring general, vocational or professional education in formal educational system (officially recognised and organisationally ensured education), i.e. in institutions of general education, vocational education or higher education. Students studying in military educational institutions were recorded as being in conscripts. Persons on academic leave were classified according to the activity they were engaged with during the leave.
2. **Pensioner** - person who had been granted an old-age pension, survivor's pension, disablement or public pension.
3. **Homemaker** - person who was not seeking work and was not ready to start work within two weeks and was engaged in unpaid household duties and/or took care of the children and other members of the family, for example, persons on parental leave (persons on pregnancy and childbirth

leave were not included) or other persons responsible for the care of the household.

4. Not working for other reasons - persons not working for other reasons, including:
  - a. persons in police or penal institutions, including police houses of detention, detention cells, prisons, pre-trial imprisonment or refugee camps;
  - b. persons who saw no reason for working or who lived off their property (rentiers);
  - c. persons who would have wanted to work but who for some reason had given up seeking work;
  - d. persons who were not seeking work and who were not ready to start work within the next two weeks as he/she attended training courses; and
  - e. persons not working due to disability or long-term illness.

Note that age 10 years were included since the prevailing poverty situation have made some minors contributors in augmentation of family income, they are therefore engaged in some form of street trading or vocational business. Areas where these situations are prevalent are indicated with a square black dot on Nigeria population map, the area with circle dot is relatively less prevalent and white small circle is scantily populated areas; this may be due to rural-urban migration.



**Fig. 1: The dotted areas are the areas densely populated with economically active population**

These subsets of the population are tabulated in table 1 of the Appendix. To the best of our knowledge no probability distribution has been provided or proposed in the present literature for this important class of population, and thus, we found the extended skew normal distribution as somewhat a pioneer model in this area. Owing to important role these classes' plays in planning, budgeting and controlling, It is essential to have a model at hand that can give the best fit to the data, when compared with the popular model that assumptions regularly made. The present paper examines a direction to the above broad based problem, namely the possibility of extending some of the classical hidden truncation method of skewing original symmetric distribution for modelling economically active population. The specific objectives are to;

present extended skew normal distribution, which was derived using the hidden truncation method of Arnold and Beaver (2000) and some of its properties that make it suitable for modelling; examined the inference from the parameters of this distribution, in which we used numerical maximum likelihood approach; and also to examine the potential application of this distribution in modelling such data as economically active population data using Nigeria economically active population data as a case study. As earlier stated the economically active data is skewed in nature as shown in the histogram in **Fig. 1** in the Appendix with fitted extended skew normal distribution.

#### **4. Numerical Results and Discussion.**

One of the methods used to test the goodness of fit in Statistical modelling is Maximum Likelihood Estimation (MLE) method. This method is chosen because it has many optimal properties in estimation: sufficiency; consistency; efficiency; and parameterization invariance which are rarely found in other approaches for detail on MLE see Spanos (1999) and Casella and Berger (2002).

Since the economically active population data are all available in grouped data. Parameters estimation have been carried out by numerically maximizing the likelihood function, by differentiating the log-likelihood function (5) with respect to each parameter  $\theta=(\mu,\sigma,\lambda,\delta)$  in

$$\log L(x | \mu, \sigma, \lambda, \delta) = -\frac{n}{2} \log \frac{\pi \sigma^2}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + \sum_{i=1}^n \log \Phi \left[ \frac{\lambda(x_i - \mu)}{\sqrt{\sigma^2 + \delta(x_i - \mu)^2}} \right] \quad (5)$$

the function and equating to zero. Since the close form of the equation cannot be easily derived we employed numerical approach by employed the R software programme the statistical computing environment (Ihaka & Gentleman, 1996), which was further supplemented with the package called "library sn" downloaded into R file from the site <http://cran.r-project.org/> to analyze the data, our data could be characterized by age-dependence structure. In fact  $x_i$  is the response of the case  $i$ .

The results are summarized in Table 3 for normal, skew normal and extended skew normal. For the case of skew normal we used the table provided by Gupta

and Chen (2001). We also considered fitting the normal distribution and compare with other distribution because, the normal distribution is the limiting or boundary case for many parametric families of probability densities, therefore, in practical statistics, the normal family is quite naturally perceived as the “central” form of a range of densities. The data used were extracted from the database of the National Bureau of Statistics a subsidiary of National Population Commission.

As earlier mentioned after removing the below 10 years cases, the economically active individual age have been categorized into a set of 7 intervals, and frequency counts obtained. Table 2 shows the summary statistical analysis for the data in Table 1.

## **5. Conclusion**

The minus log-likelihood function of the parameters in the estimated models show that there is no difference between the normal and the skew normal models. A better fit is obtained with the extended version which has additional parameters  $\lambda$  and  $\delta$  with lesser minus log-likelihood when compared with normal and skew normal model. The location and the scale parameters for the normal and skew normal are the same. This shows that the extended type of the skew normal distribution could be adopted as a model for economically active population because it's fit the data satisfactorily, when compare to other commonly used models. Only that its parameters could not be explicitly expressed and its cumulative distribution function is not in closed form. The probability plot of the histogram and the fitted extended skew normal model are given in fig. 2 to further illustrate the applicability of the model. Some Mathematical and Statistical properties that make the extended types of the skew normal suitable may still be explored in future research, also further area of applications because of its generalized nature over normal and skew normal densities.

## **6. Acknowledgements**

I will like to express my sincere gratitude to National Bureau of Statistics for given me access to their database and also to the two anonymous referees, whose intense criticism makes the paper better.

The research was part of work carried out at Dipartimento di Scienze Statistiche Università Degli Studi di Padova, Italy. When the author was on COIMBRA fellowship research visit.

## **References**

1. Adcock, C. (2004). Capital asset pricing in UK stocks under the multivariate skew-normal distribution. In Genton, M. G., editor, *Skew-elliptical distributions and their applications: A journey beyond normality*, chapter 11, pages 191–204. Chapman & Hall/CRC.
2. Arnold, B. C. & Beaver, R. J. (2000). Hidden truncation models. *Sankhya*, ser. A. 62: 22–35.

3. Azzalini, A. (1985). A Class of Distribution which includes the Normal Ones, *Scand. J. Statist.* 12: 171-178.
4. Azzalini, A. and Capitanio (1999). Statistical applications of the multivariate skew-normal distributions. *J.R. Stat. Soc. series B* 65: 367-389.
5. Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* 83: 715-726.
6. Bolfarine, H. & Lachos, V. (2006). Skew binary regression with measurement errors. *Statistics* 40: 485–494.
7. Cassela, G., and Berger, R.L. (2002). *Statistical Inference* (2<sup>nd</sup> ed.). Pacific Grove, CA: Duxberry.
8. Chen, J. T., Gupta, A. K., & Troskie, C. G. (2003). The distribution of stock returns when the market is up. *comms-tm* 32: 1541–1558.
9. De Luca, G., Genton, M. G., & Loperfido, N. (2005). A multivariate skew-GARCH model. *Advances in Econometrics*, 20: 33–57.
10. Ellison, B.E. (1964). Two theorems for inferences about the normal distribution with applications in acceptance sampling. *Journal of the American Statistical association*, 59: 89-95.
11. Gupta, A.K. and Chen, T. (2001). Goodness-of-fit tests for the skew-normal distribution. *Commun. Statist.-Simula.* 30(4), 907-930.
12. Gupta, R. C. & Gupta, R.D. (2004). Generalized skew normal model. *Test* 13: 501–524.
13. Ihaka, R. and Gentleman, R. (1996). A language for data analysis and graphics. *Journal of Computational and Graphical Statist.*, 5: 299-314.
14. Spanos, A. (1999). *Probability theory and statistical inference*. Cambridge, UK: Cambridge University Press.
15. Zack, S. (1981). *Parametric statistical inference*. Pergamon Press. Oxford.

## Appendix

**Table 1: Economically active population grouped by their ages**

Age midpoints ( $X_i$ )	Frequencies distribution
14.5	2390641
24.5	7111678
34.5	6833138
44.5	4569953
54.5	2696806
64.5	1730112
74.5	1292598

**Sources:** FOS/ILO/SIMPOC, National Modular Child Labour Survey, NBS, Abuja Nigeria.

**Table 2: Summary for Table 1**

Minimum	Maximum	Mean	Median	Mode	Lower quartile	Upper quartile
9.5	89.5	37.7	35.1	28.9	28.5	47.5

**Table 3: Maximum likelihood estimates of the parameters of each distribution considered as possible model for the economically active data.**

Parameters estimates	Normal	Skew-Normal	Extended skew Normal
Mean	37.667	37.6564	33.400
$\sigma^2$	246.416	246.416	183.214
$\Lambda$	Nil	0.01	0.01
$\Delta$	Nil	Nil	24.419
Log-likelihood	-7.6009	-7.6009	-6.80941

**Note:** the log-likelihood function is multiplied by  $10^6$

**Fig. 2:** Histogram (based on probability plot) of economically active population with fitted extended skew normal distribution.