IOSR Journal Of Humanities And Social Science (IOSR-JHSS)
Volume 21, Issue 4, Ver. 04 (Apr. 2016) PP 35-40
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Tweedie Modelling For the Determinants of Child Mortality

S. Santhana Lakshmi[1],R. Geetha[2]
[1]*Research Scholar Department of Statistics S.D.N.B Vaishnava College for Women Chrompet Chennai-44, Tamil Nadu, India.*
[2]*Associate Professor Department of Statistics S.D.N.B.Vaishnava College for Women Chrompet Chennai-44, Tamil Nadu, India.*

***Abstract**: Socio–economic, biological and demographic factors may impact the IMR and it is hoped that it can be explained in terms of Mother's age, Birth order, Age at First marriage, religion, gender, residence, Birth weight of infant etc. So far research studies have used GLM, Joint GLM based on log Gaussian and Gamma distribution (McCullagh & Nelder (1989), Lee & Nelder (1998, 2003)) to identify the risk factors of child mortality. Tweedie distribution (Tweedie(1984),Jorgensen (1997) has the potential to model both discrete and continuous processes. This paper attempts to model infant survival time as a function of household and maternal predictors using NFHS-3 data by taking into account the non-zero probability of zero occurrence of the response variable 'infant age at death'.It is found that Tweedie model performs better than Gamma model based on Akaike Information Criterion (AIC).*

***Keywords:**Tweedie distribution, Gamma model, child mortality, AIC, NFHS –3.*

I. INTRODUCTION

India has predominantly huge growth in population in the last 50 years and has become ahighly populated country all over the world. Infant Mortality rate(IMR) is the degree to calculate the death rate of children less than 1 year and is a sensitive factor to determine the country's welfare and development. Developed nations are largely successful in managing the IMR rate, but under developed countries still fail to face any decline in IMR and it becomes a greater challenge to society on the public health system.India with 43 infantdeaths per 1000 during 2014 has shown a significant decline in IMR in 2015 with 32 infant deaths per 1000. So far research studies in IMR have developed a set of statistical models to analyse the effect of risk factors on infant survival time (Firth 1988, McCullagh & Nelder, 1989).Lee & Nelder (1998, 2003) have proposed the use of Joint generalised linear model based on Log- Gaussian and Gamma distributionto allow for structured dispersion in the data.Das (2011) has used National Family Health Survey (NFHS-2) data to identify the significant factors among maternal, proximate and household determinants of child mortality.NFHS survey provides data on family welfare, maternal,marriage, fertility, family planning, reproductive health, child health, nutrition and HIV/AIDS. For each live birth age, sex and survival status of the child is noted and for dead children, age at death is recorded. However, these models fail to capture the non-zero probability of zero occurrence of an event. Since infant survival time involves both discrete (survival time = 0) and continuous part (survival time >0), a model which combines both circumstances namely the Tweedie model is more appropriate.Tweedie family has been appliedin various fields like actuarial science,economics, telecommunication, medicine and ecology, Perry (1981),Robert Gilchrist and Denise Drinkwater (1999), Dunn (2004), Dunn & Smyth, (2005), Robert Gilchrist (2007),and Hiroshi Shono (2008).With a view to study the impact of the maternal factors Mother's age, Parity, age at first marriage and birth weight on infant survival time, two models based on Gamma and Tweedie distribution is considered in this paper. Age at death of infant (in months) is taken as a measure of child mortality. The rest of the paper is organized as follows. In section 2 Tweedie distribution and generation of Tweedie random variables is discussed. Section 3 presents the methods and findings of NFHS dataand results are given in Section 4.

II. TWEEDIE DISTRIBUTION

Tweedie distribution is a mixture of degenerate distribution and belongs to a class of exponential family distribution for which $V(\mu) = \mu^p$ and$V(y) = \varphi\mu^p$ for some p. This family includes normal (p=0), Poisson (p=1), gamma (p=2), and inverse Gaussian (p=3).This family includes discrete, continuous as well as mixed densities. Tweedie model is closely connected to the dispersion exponential model (Jorgensen(1987, 1992), Smyth (1996)), which has the form

$$f(\mathrm{y};\theta,\varphi) = \mathrm{a}(\mathrm{y},\varphi)\exp(y\theta - k(\theta)/\varphi) \qquad (1)$$

where $\theta = \mu^{1-p}/1-p,\ k(\theta) = \mu^{2-p}/2-p$

For exponential family of distribution,

$$\mu = E[Y] = dk_p(\theta)/d\theta \text{ and } , Var(Y) = d^2k_p(\theta)/d(\theta^2) = \mu^p.$$

Then equation (1) can be written as

$$f(y;\theta,\varphi) = a(y,\varphi)\exp\left[1/_{\varphi}\left(\frac{y\mu}{1-\rho}^{1-\rho} - \frac{\mu}{2-\rho}^{2-\rho}\right)\right] \quad (2)$$

2.1Generation of Tweedie Random Variable:

Let N, $X_1, X_2, \ldots\ldots, X_r$ be a sequence of independent random variables such that N is Poisson distributed and X_r, r=1, 2… N is identically distributed. The response variable Yis exact zeros with non-zero- probability, or,otherwise, positive and continuous density of Y. The standard method of analysing such data is using a special form of Tweedie compound Poisson distribution in which the response variable Y is assumed to generate a sum of individual random variables with positive values. That is$Y = \sum_i^N X_i$, where N is the number of event follows a discrete count distribution and $X_i$is the magnitude outcome in the i^{th} event. When N=0, we have Y=0, this distribution has a probability mass at zero. The mean and variance relation is specified to be$Var(Y) = \varphi\mu^p$, where p is the power parameter ($1<p<2$).

If N∼ poisson (λ) and X_i∼gamma (α, θ), are independently and identically distributed gamma random variables with shape parameter α and the scale parameter θ then the parameter λ α and θ are related to the natural parameters $\mu, \varphi, and\ p$ of the Tweedie distribution given in (1) as

$$\lambda = \mu^{2-p}/\emptyset(2-p)$$
$$\alpha = 2-p/1-p$$
$$\theta = \emptyset(1-p)\mu^{1-p}$$

The mean of the Tweedie distribution is positive for p>1. It is well defined for all p expect $0<p<1$. Our particular interest is $1<p<2$.

TheTweedie $T_p(\mu,\sigma^2)$random variables Y requires N to have mean $\lambda K_p(\theta)$, for $\lambda = 1/\sigma^2>0$ where$k_p(\theta) = \frac{1}{2-p}[\theta(1-p)]^{(2-p/(1-p)}$, with $1<p<2$, and$X_r$is distributed according to the gamma distribution with shape parameter rk and $E[X_r] = {-rk}/_{\lambda\theta}$.

Then P(Y=0) = P(N=0) = exp ($-\lambda k_p(\theta)$), and the density of Y for Y>0 is given by $f_y(y) = \sum_{r=1}^{\infty} P(N=r) f_{X_r}(y)$ where $f_{X_r}(y)$ is the density ofX_r.

III. METHODS

A. Data

The Study is based on secondary data from National Family Health Survey (NFHS -3, 2005-06).The survey is conducted every 6 years with respondents being ever married women in the age 15- 49 groups. In this paper the data is extracted from 10178 women from Rajasthan to model IMR based on the Tweedie and Gamma distribution with seven predictors of child mortality-Place of residence, religion, and infant gender are the household factors and Mother's age, Birth order, Birth weight, and Age at first marriage are the maternal factors.

Data sources-NFHS-3(2005-06).

STATE-WISE DISTRIBUTION OF WOMEN (15 - 49) years

According to NFHS -3, Uttarpradesh has highest percentage of 13% ever married woman, Maharastra and Madhyapradesh holds 6%;AndhraPradesh and West Bengal holds 5%; Karnataka, Bihar, Rajasthan,Tamil Nadu, Orissa holds 4%; Chhattisgarh, Nagaland, Manipur, Gujarat, Assam, Jharkhand, Punjab, Delhi holds 3%; Jammu and Kashmir,Uttaranchal,Haryana,Himachal Pradesh, Kerala,Goa, Meghalaya holds 2% and Arunachal Pradesh,Tripura,Mizoram,Sikkim state has least percentage of 1%.

B. Description of covariates and levels:

1. Dependent Variable:

The dependent variable for study is the infant survival time. (Age at death is measured in months).

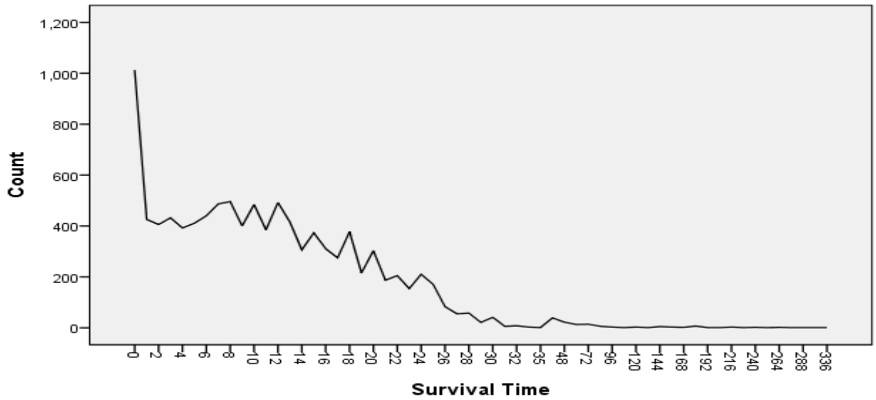
2. Independent Variable:

Mothers Age (M Age), Birth Order (BORD), Birth Weight, Age at First marriage, Place of residence , Religion and Gender are taken as independent variables for our illustration. The covariates are categorised in Table 1. Fig 1 shows the distribution of infant mortality in the state Rajasthan.

Table 1Categorisation of Variables in the Analysis

Domain/Variable Name		Description
Household/Community Factors	Place of residence	Rural -0,Urban-1
	Gender	Male -0 , Female -1
	Religion	Hindu -1, Muslim 2,others-3
Maternal Factors	Age at First Marriage	In Years
	Birth Order	Parity
	Birthweight	In Kilograms
	Mage (Mother's age)	In years
Dependent Variable	Infant survival time	Age in Months

Fi**g (1): Infant mortality distribution in Rajasthan, India: 2005 – 2006**



3.1 FINDINGS

A. Descriptive Statistics:

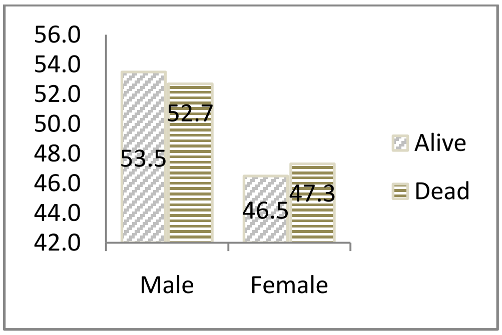
In this study, a vast majority of people 71% live in urban areas. The religion wise distribution is 87 % Hindu, 12% Muslim and others 1 percent. Among infants, 53.4 % are male and46.6 % are female. Table 2 gives the mean and standard deviations for all maternal factors.

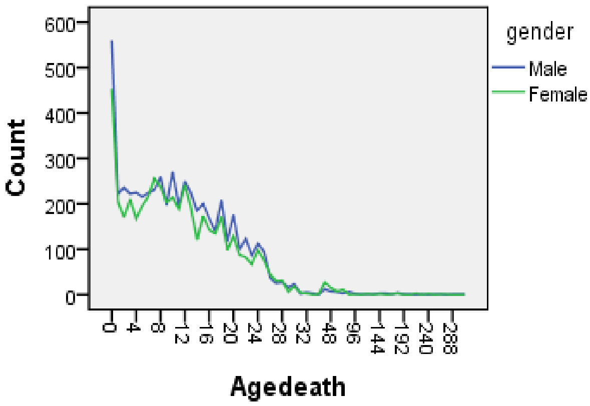
Table2: Means and Standard Deviations for all Maternal Factors

Variable Name	Mean	Standard Deviation
Mage in years	35.15	7.90
Birth Order	2.90	1.89
Birth Weight	8.34	3.07
Age at First Marriage in years	15.87	2.7

Fig (2) shows the sex wise distribution of children according to their survival status. We find that the survival proportion is same across gender in the state of Rajasthan. Fig (3) shows gender wise distribution of child mortality in terms of survival time in months.

Fig (2) Percentage distribution of children dead/aliveFig (3) Gender wise Distribution of child mortality





The range of mean infant survival time(in months) of males is 11.59±14.86 and female is 11.55±13.79. The overall mean ± sd of child mortality is 11.57 ± 14.37 in months. Fig 4 shows sex wise distribution of infants whose survival status is "dead".The range of infant survival time dead (in months) of male is 12.52±35.94 and female 14.20±33.41. The overall mean ± sd of child mortality of infant survival time dead in months is 13.52 ± 35.49 in months.

Fig (4) Distribution of child insurvival status 'dead'

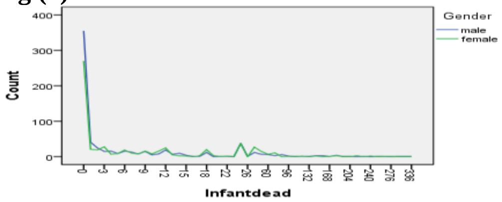


Table 3: Distribution of Infant Survival (Status: dead)

Gender	Male		Female		Total
	Number	%	Number	%	Number
Age at death (>0)	205	52.8	183	47.2	388
Age at death Zero	355	56.8	270	43.2	625
Total	560	55.3	453	43.7	1013

Table 3 represents thesexwise distribution of infants whose age at death is "Zero"and "greater than zero". It is evident that there is a non zero probability of 0.62 for the event "infant age at death is zero". It is 0.63 for males and 0.6 for females respectively. Fig 5 shows the distribution of infant whose survival status is "dead". Table 4 represent the results of Generalised linear gamma model and Tweedie model with seven predictors.

Fig 5:Infant survival time (Status - dead)

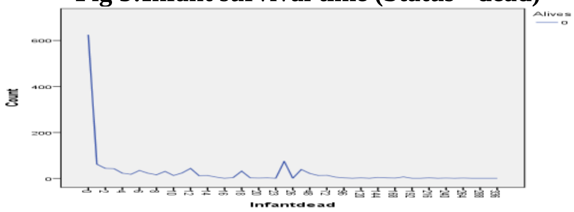


Table4.ResultsfromGamma and Tweedie model.

COVARIATES	Gamma Model				Tweedie Model			
Parameter	**Estimate B**	**Std. Error**	**Sig.**	**Exp(B)**	**Estimate B**	**Std. Error**	**Sig.**	**Exp(B)**
(Intercept)	.744	.1288	.000	2.104	-.731	.1660	.000	0.481
[Gender=1]	-.060	.0304	**.047**	0.942	.006	.0391	.875	1.006
[Religion=1]	.065	.1520	.670	1.067	.070	.1921	.717	1.073
[Religion=2]	.016	.0461	.733	1.016	-.023	.0590	.693	0.977
[Residence=1]	.065	.0392	.097	1.067	.079	.0510	.119	1.082
BORD	.049	.0055	**0.00**	1.050	.064	.0065	**0.00**	1.066
Mage	-.093	.0136	**0.00**	0.911	-.120	.0169	**0.00**	0.887
Age at first marriage	-.054	.0071	**0.00**	0.947	-.059	.0086	**0.00**	0.943
Birthweight	.018	.0056	**.002**	1.018	.017	.0073	**.025**	1.017
(Scale)	.352[b]	.0120			.536[b]	.0142		
AIC	**5604.468**				**2577.983**			

1. BORD, Mage, age at first marriage and Birth weight,is statistically significant ($p<0.05$) in both models.

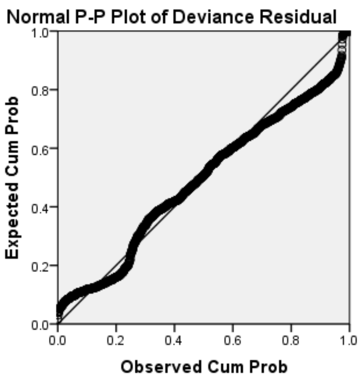
2. BORD, Birth weight is statistically significant ($p<0.05$) and it is positively associated with infant survival time in both models.
3. Mage and age at first marriage is statistically significant ($p<0.05$) and it is negatively associated with infant survival time in both models.
4. Gender is statistically significant ($p<0.05$) and it is negatively associated with infant survival time in Gamma model only.
5. The power parameter of Tweedie model is estimated as 0.536.

Gamma model:
Y=exp (.744 -.060*male+.065*Religion (1) +.016* Religion (2) +.065*Urban+.049*BORD-.093*Mage-.054*Age at First Marriage+.018*Birthweight)

Tweedie model:
log (y+.01) = exp (-.731+.006*male+.07* Religion (1) -.023* Religion (2) +.079*Urban+.064*BORD-.12*Mage-.059*Age at First Marriage+.017*Birthweight)

As the measure of goodness of fit, the AIC value for Tweedie model is 2577.983 which is very less than the Gamma model, though the standard error of the parameter are slightly smaller in Gamma model. We examined the goodness of the model fit based on graphical analysis, by plottingthe deviance residuals as shown inFig 6(a) and Fig 6(b).The pp – plots does not reveal any lack of fit. Thus we conclude that Tweedie model is more efficient than Gamma model.

PP PLOT OF THE RESIDUALS



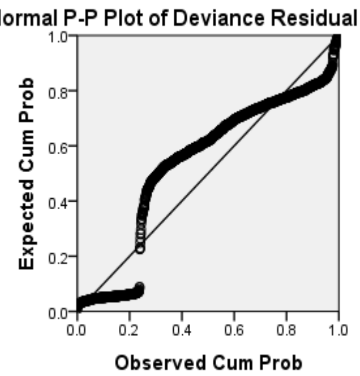


Fig. 6 (a) Fig.6 (b)

IV. CONCLUSION

Children are dynamic resource to the nation's present and its future. Ensuring the survival and wellbeing of children is a concern of families, communities and nations throughout the world.Infant mortality is a sensitive indicator of a country's growth and development.NFHS provides data on several predictors of child mortality like mother's age, birth order, age at marriage, birthweight, religion, gender, residence etc.In this study on 10178 ever married women in age15-49 yearfrom Rajasthan state in India,the response variable 'infant age at death' is either zero with non -zero probability or,otherwise has positive and continuous density. Tweedie distribution and Gamma models were fitted to the data.

Our findings reveal that in modelling IMR the Birth order, Age at first marriage, Birth weight and mother's age are statistically significant in both Tweedie and Gamma models. Gender of infant is a significant predictor only in Gamma model. Mother's age and age at marriage are negatively associated with child survival in both models. Birth order and birth weight are positively associated with infant survival time in both models.

Tweedie model is much more efficient than Gamma model based on AIC values.This paper demonstrates the efficiency of Tweedie distribution in modelling the determinants of child mortality when there is a non-zero density for positive values,but with a point mass at the origin. Further studies using NFHS data for all states can be carried out to identify the risk factors on IMR.

REFERENCES

[1]. Das RN, Lee Y. Log-normal versus gamma models for analysing data from quality improvement experiments. Quality Engineering 2009; 21(1): 79-87.

[2]. Gilchrist, R. (2000). Regression models for data with a non – probability of a zero responses. Communications in statistics Theory and Methods 29(9).
[3]. Gilchrist, R. and Drinkwater, D. (1999).Fitting Tweedie models to data with probability of zero responses.
[4]. Hiroshi Shono. Application of the Tweedie distribution to zero-catch data in CPUE analysis.Fisheries Research 93 (2008) 154–162.
[5]. Rabindranath Das etal (2011) Infant Mortality in India: Evaluating Log-Gaussian and Gamma Distributions, The open Demography Journal, vol.4, 34-41.
[6]. Jorgensen, B. J., the Theory of Dispersion Models,Chapman & Hall, 1997.
[7]. Jorgensen, B., and M. C. P. De Souza, "Fitting Tweedie's Compound Poisson Model to Insurance Claims Data," ScandinavianActuarial Journal 1, 1994, pp. 69–93.
[8]. Jorgensen, B. (1987) ExponentialDispersion Models, J.R. Statist. Soc. B 49,127-162.
[9]. Dunn, P.K. and Smyth, G.K (1996).Randomized Quantile Residuals. J.Compute.Graph. Statist. 5, 236-244.
[10]. Dunn, P.K. and Smyth, G.K (2005)."Series Evaluation of Tweedie Exponential Dispersion Models Densities". Statistics and Computing, 15, 267-280.
[11]. Tweedie MCK (1984). An index which distinguishes between some important exponential families. Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Indian Statistical Institute, Calcutta, pp. 579-604.
[12]. Smyth, G. K., and Jorgensen, B. 2002. Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. ASTIN Bulletin 32: 143{157.