Research Journal of Mathematical and Statistical Sciences ______________________________ E-ISSN 2320–6047
Vol. **4(2),** 1-5, March **(2016)** *Res. J. Mathematical and Statistical Sci.*

Joint GLM of Mean and Dispersion for the Determinants of Child Mortality

S. Santhana Lakshmi and R. Geetha*
Department of Statistics, S.D.N.B Vaishnav College for Women chromepet, Chennai-44, Tamil Nadu, India
hayageeth@hotmail.com

Available online at: www.isca.in, www.isca.me
Received 21st January 2016, revised 12th February 2016, accepted 3rd March 2016

Abstract

Joint modelling of the mean and dispersion parameters as a function of explanatory variables is widely used for positive data analysis. Log - Gaussian or Gamma models for constant variance and Joint GLM for non- constant variance (Mc Cullagh and Nelder, 1989) are reported in literature. The aim of this paper is to apply this technique to identify the effects of some determinants in child mortality using data from NFHS-3 (2005-06). It is found that in modelling structured dispersion, Log - Gaussian model is better than Gamma model based on Akaike Information Criterion (AIC).

Keywords: Joint GLM, Log- Gaussian, Gamma model, Child mortality, AIC, NFHS – 3 (2005-06).

Introduction

Infant Mortality Rate is number of deaths of infant less than one year old per 1000 live births in a given year. This rate is an indicator of the level of health that reflects a country's growth and development. Infant deaths are classified as neonatal deaths and post-neonatal deaths based on infant age at death. In India, IMR rate shows significant decline in the past few decades and in 2014 it is ranked 45th in the world with 43 infant deaths per 1,000 live population[1]. But still the magnitude of IMR pose a significant challenge to the society and to the public health system in more developed countries[2]. India expects to further reduce child mortality with increased allocation of funds and stepping up investments in health system and improve maternal health this year. There has been a significant decline in under-five, neo-natal and maternal mortality rates during 2013 due to innovative health interventions that have led to increase in the life expectancy and rates of immunisation of children. With this the current rate of decline, India is likely to achieve its target of reducing the infant mortality rate to less than 32 per 1,000 live births by end of 2015. The focus of this paper is to examine the possible risk factors of childhood mortality in Indian states and identify best statistical model by joint modelling of mean and dispersion parameters as functions of explanatory variables. The rest of the paper is organized as follows. In section 2 various factors affecting IMR and Joint GLM models is discussed. Section 3 applies the model to NFHS -3 data and the results and findings are presented in section 4.

Causes of IMR

Mortality is often used as a barometer of welfare. The factors affecting IMR is mainly due to biological, behavioural and environmental conditions. Delhi has highest infant mortality rate among other metros in India. Chennai recorded the lowest IMR of 15 per 1,000 live births and in Mumbai and Kolkata it was 20 per 1,000 live births during 2014[3]. Ninety per cent of these deaths occur due to easily preventable diseases like pneumonia and diarrhoea[4]. A different medical and public health approaches are needed to marks this transition from infections to neonatal conditions and premature births.

Biological factor includes mother's age, order of birth, time interval, and birth weight, multiple births or pre-mature birth. The characteristics may include maternal, fetal and obstetrical factors. Demographic variables like mean age of marriage, mean age of cohabitation, age at first birth marriage below age eighteen, may have an impact on infant survival time[5].

Environmental factors include social, economic and cultural factors like religion, Place of residence lack of medicines, infections, premature births, behavioural changes in fertility, nutritional status, usage of health services, complications during delivery, perinatal asphyxia, birth injuries, communicable diseases, congestion, living condition, food security, feeding practices, increase in urbanization, per capita income, monthly per capita consumption expenditure and household facilities like electrification, sanitation. Behavioural characteristics such as cigarette, alcohol, drugs, region of living may cause the preterm birth, restriction in fetal growth, low birth weight and low birth height.

Alcohol and Stress are considered as a high risk factor and psychological reaction to stress might consequently affect the child's characteristics at birth stage. Breastfeeding affects the health and nutritional status of the mother and child. The influences of parental education on infant and child health from the education level of parents and their work status each have independent effects upon child survival in developing countries.

Background: Various researchers of India's Infant Mortality Rate discussed the risk factors in the order of: proximate, maternal, and household /community factor. Proximate factors include items that require medical care and non – medical care during the antenatal, birth, and postnatal period. Maternal factors refer to items such as mother age, birth order, birth interval, and age at first marriage. A household/community factor refers to items such as sanitation, water supply and household cleanliness. These factors significantly increase IMR without any development in socio-economic status.

It is of interest to set up statistical models to analyse the effect of risk factors on infant survival time. In general, regression analyses of positive observations are based on log normal and gamma models[6-8]. Usually positive observations are non-normal and the variance may or may not be constant. In regression modelling to stabilize variance appropriate transformation of the response variable is used. In practice variance may not be stabilized always[9].

Under this allow for structured dispersion Lee and Nelder have proposed the use of Joint Generalized Linear Model (JGLM) [10]. Das has used NFHS-2 data for infant survival time in Bihar and infers that Joint Log Gaussian Model (LGM) fit is more efficient than Joint Gamma Model (GM) fit based on AIC criterion. In the former model all the maternal, proximate and household/community effects are significant whereas mother's age is not significant in joint GM model[6]. Das and Lee in the analysis of resistivity, a positive characteristic in quality industry, Das to identify the factors of mother's life style characteristics which have statistical significant effects on her neonate birth weight, efficiency of modelling of structured dispersion[10-11]. Gurprit Grover has used gamma GLM for estimation of survival function of Diabetic Nephropathy patients[9].

With a view to study the impact of the maternal factors Parity, age at first marriage and birth weight on infant survival time, two models based on lognormal and gamma distribution is considered in this paper. Age at death of infant (in months) is taken as a measure of child mortality.

Models: The general structure of JGLM model is described in Model 1 and Model 2.

Model 1: Multiplicative model (Gamma)
Let the response variable $Y_i = \mu_i$ (i = 1, 2, … n)
where: $\eta_i = \log \mu_i = x_i^t\beta = \beta_0 + x_{i1}\beta_1 + \ldots\ldots\ldots + x_{ip}\beta_p$ and $\{\epsilon_i\}$s are i.i.d with $E(\epsilon_i) = 1$.then

The JGLM for the multiplicative model is
$E(Y_i) = \mu_{Yi}$ and $Var(Y_i) = \sigma^2_{Yi}\mu^2_{Yi}$, where $\eta_i = (\log(\mu_{Yi}) = x_i^t\beta_Y$ and $\xi_i = \log(\sigma^2_{Yi}) = g_i^t\gamma_Y$, g_i is the row vectors of the model matrix used in dispersion model.

Model 2: Log Normal
If $Z_i = \log(Y_i)$ and if Y_i follows a log normal distribution i.e. $Z_i \sim N(\mu_{zi}, \sigma 2)$

$\mu_{Yi} = E(\exp Z_i) = \exp(\mu_{zi} + \sigma^2/2) \neq \exp(\mu_{zi})$
The JGLM for the log normal model is
$E(Z_i) = \mu_{zi}$ and $Var(Z_i) = \sigma^2_{zi}$, where $\mu_{zi} = x_i^t\beta_z$ and $\xi_i = \log(\sigma^2{}_{zi}) = g_i^t\gamma_z$

Method of estimation: ML Estimators for β's and Restricted ML Estimators for γ's are obtained using two interconnected iterative weighted least squares (Lee et.al, 2006).

Evaluation criteria: i. Akaike information criterion (AICs) for the best fitting model, ii. standard errors of estimates and iii. Graphical analysis.

Methods: A. Data: The study is based on secondary data from National Family Health Survey (NFHS – 3,2005-06). This house hold survey is conducted throughout India by International Institute for Population Sciences (IIPS), Mumbai, under the Ministry of Health and Family Welfare (MOHFW), Government of India[12].

The survey is conducted every 6 years with respondents being ever married women in the age groups 15- 49. It provides state and national-level information on fertility, family planning, infant and child mortality, maternal and child health, nutrition of woman and children, etc.
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%, Andhra Pradesh 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%.

In our analysis 10178 ever married women in 15-49 age groups from Rajasthan states is considered to demonstrate the efficiency of joint modelling of mean and dispersion in Infant survival time.

Description of covariates and levels:
Dependent Variable: The dependent variable for study is the infant survival time. (Age at death is measured in months). Figure-1 shows the distribution of Infant mortality in Rajasthan.

Independent Variable: The maternal factors Mothers Age (Mage), Birth Order (BORD), Age at First marriage, Birth weight and the household/community factors Place of residence, Gender, Religion are taken as independent variables. The covariates are categorised as in Table-1 below.

Table-1
Categorisation of Variables in the Analysis

Domain/Variable Name	**Description**
Household/Community	
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
Mage (mother's age)	In years
Birth weight	In Kilograms
Dependent Variable	
Infant survival time	Age at death in Months

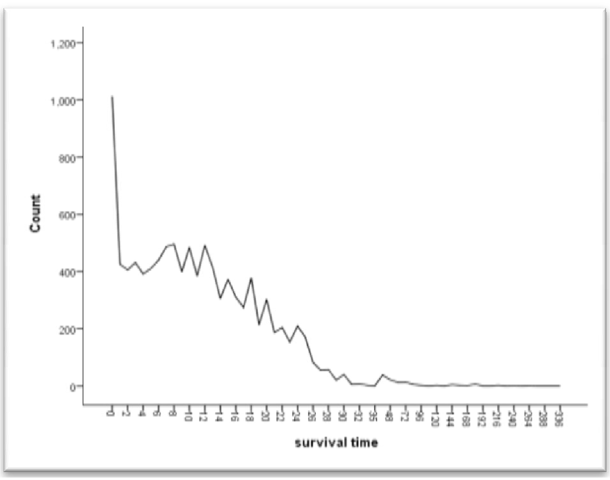


Figure-1
Infant mortality distribution in Rajasthan, India: 2005 – 2006

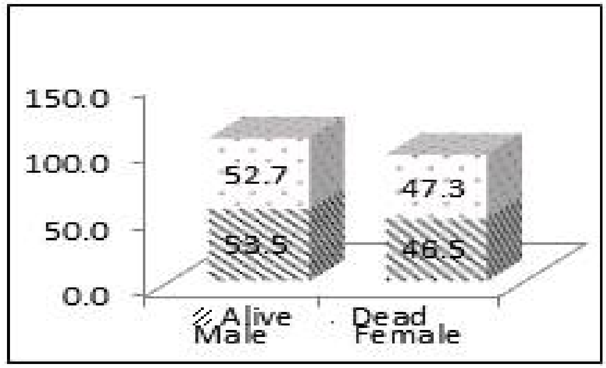


Figure-2
Percentage distribution of children dead/alive

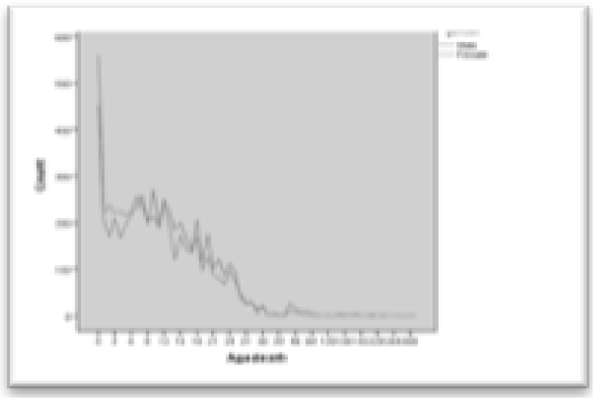


Figure-3
Gender wise Distribution of child Mortality

Findings

Descriptive Statistics: The data for state of Rajasthan suggest that the majority of the people live in urban areas (71 %), and the majority of people are Hindus (87 %). The sex wise percentage distributions of infant are 53.4% male and 46.6% female. Figure-2 shows percentage distributions of infant death. Figure-3 shows gender wise distribution of child mortality. Table-2 gives the mean and standard deviation for all maternal factors.

The range of mean infant survival time (in months) of males is 11.59 ± 14.86 and female is 11.35 ± 13.79. The overall mean ± sd of child mortality is 11.57 ± 14.37 in months.

Table-2
Means and Standard Deviations for all Maternal Factors

Variable Name	**Mean**	**Standard Deviation**
Maternal Factors		
Mage in years	35.15	7.904
Birth Order	2.90	1.894
Birth Weight(in kg)	8.34	3.07
Age at First Marriage in years	15.87	2.7

Fitting Log – Gaussian and Gamma Models: Data obtained from NFHS- 3 on ever married woman in the state of Rajasthan is used in this study. The model for mean was first fitted, and the variance model was then fitted based on the corrected or uncorrected squared residuals from the mean model fit. The

Joint GLM results from the Log – Gaussian Model and Gamma Model are presented in Table-3.

We find that in the mean model, i. Mage, BORD, Age at first marriage and Birth weight is statistically significant ($p<0.05$) in both models. ii. Gender of infant is a significant predictor only in Gamma model.

In the dispersion model: Mage, BORD and Age at first marriage are statistically significant ($p<0.05$) in both models whereas gender is statistically significant ($p<0.05$) in joint LGM model.

Model-1: Y=exp (.744+.065*Urban+.065*Religion (1) +.016* Religion (2)-.006*Male+0.049*Mage-0.093*BORD -.054*Age at First Marriage+0.018*Birth weight)

σ^2 = exp (.352+.061*Urban+.123* Religion (1) -.023* Religion (2)+.013*Male+0.075*Mage-0.144*BORD-.063*Age at First Marriage+0.008*Birth weight)

Model-2: Y= exp (-.0317-.166*Urban-.046* Religion (1)-.057* Religion (2) -.041*Male-.11*Mage+.219*BORD+.133*Age at First Marriage-0.058*Birth weight)

σ^2 = exp (.378-.014*Urban+0.075* Religion (1)+0.054* Religion (2)-.124*Male-.057*Mage+.118*BORD +.032*Age at First Marriage-.005*Birth weight)

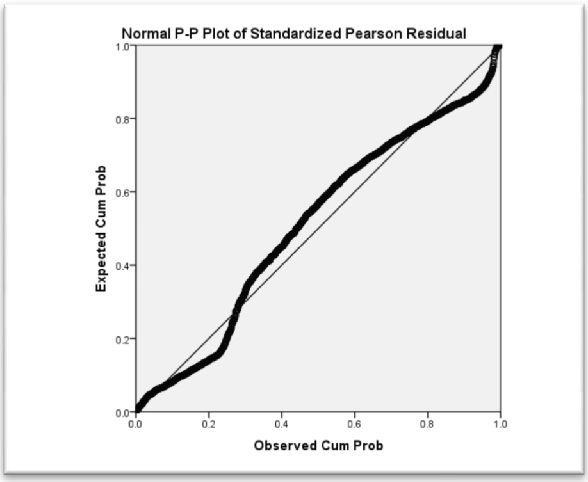
It shows that joint LGM fit is more efficient than joint GM fit based on AIC values. In both models the AIC values are reduced in mean and dispersion models and the standard errors are slightly smaller in Gamma model[4].

We examined the goodness of the model fit based on graphical analysis. In Figure-4(a) and Figure-4(b), we plot the absolute values of residuals against the fitted values for joint LGM and GM respectively. The pp – plots does not reveal any lack of fit.

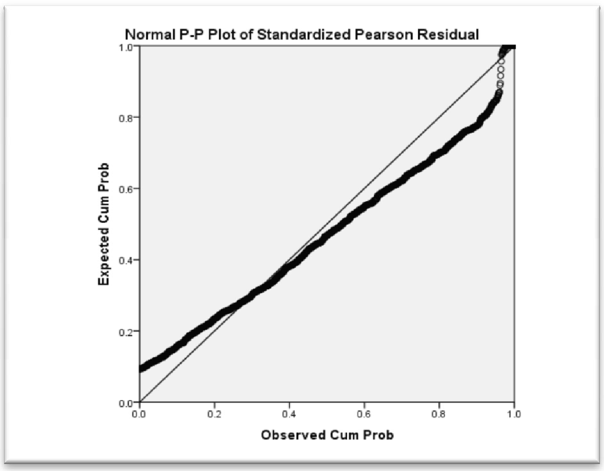
Table-3
Log – Gaussian and Gamma fit mean and constant dispersion of infant survival time

	COVARIATES	**JOINT GAMMA MODEL**				**JOINT LOG GAUSSIAN MODEL**			
	Parameter	**Estimate B**	**Std. Error**	**Sig.**	**Exp(B)**	**Estimate B**	**Std. Error**	**Sig.**	**Exp(B)**
MEAN	(Intercept)	0.744	0.1288	0	2.105	-0.317	0.3814	0.405	0.7281
	[Residence=1]	0.065	0.0392	0.097	1.067	-0.166	0.1217	0.174	0.8474
	[Religion=1]	0.065	0.152	0.67	1.067	0.046	0.412	0.911	1.047
	[Religion=2]	0.016	0.0461	0.733	1.016	0.057	0.1357	0.675	1.058
	[Gender=1]	-0.06	0.0304	0.047	0.941	-0.041	0.0925	0.659	0.9599
	Mage	0.049	0.0055	0.00	1.05	-0.11	0.0198	0.00	0.8961
	BORD	-0.093	0.0136	0.00	0.911	0.219	0.0449	0.00	1.244
	Age marriage	-0.054	0.0071	0.00	0.947	0.133	0.0243	0.00	1.142
	B weight	0.018	0.0056	0.002	1.018	-0.058	0.0169	0.001	0.9436
	(Scale)	.352[b]	0.012			1.353[b]	0.0749		
	AIC	5604.468				2067.359			
DISPERSION	(Intercept)	0.352	0.1124	0.002	1.421	0.378	0.1961	0.054	1.459
	[Residence=1]	0.061	0.0347	0.078	1.063	-0.014	0.0579	0.812	0.9863
	[Religion=1]	0.123	0.1429	0.389	1.131	0.075	0.2475	0.761	1.078
	[Religion=2]	-0.023	0.0412	0.577	0.9773	0.054	0.0685	0.434	1.055
	[Gender=1]	0.013	0.0265	0.618	1.013	-0.124	0.0451	0.006	0.883
	Mage	0.075	0.0042	0.00	1.078	-0.057	0.0098	0.00	0.9446
	BORD	-0.144	0.0114	0.00	0.865	0.118	0.0219	0.00	1.125
	Age marriage	-0.063	0.0058	0.00	0.938	0.032	0.0116	0.005	1.032
	B weight	0.008	0.0048	0.087	1.008	-0.005	0.0083	0.54	0.994
	(Scale)	.352[b]	0.0111			.563[b]	0.0218		
	AIC	3649.351				1368.015			

The fitted mean and variance model for infant survival time are



(a)



(b)

Figure-4
PP Plot of Residuals

Conclusion

Infant mortality is sensitive indicator of social-media problem that impact country's development. In this study, the focus on the determinants of child mortality include demographic factors like age at first marriage , environmental factors like place of residence, religion and biological factors like Mother's age, birth order, birthweight, and infant gender. To find the effect of these factors on infant survival time, simultaneous modelling of the mean and dispersion using joint Generalised linear models based on log Gaussian and gamma distribution is proposed.

Our findings reveal that Log Gaussian is much more efficient than Gamma model based on AIC values. In modelling mean response, Mother's age, Birth order, Age at first marriage and Birth weight is statistically significant in both models. Gender of infant is a significant predictor only in Gamma model. However joint modelling of mean and dispersion identifies infant gender as a significant predictor in joint LGM model.

This paper demonstrates the efficiency of joint GLM approach using lognormal and gamma distribution in modelling the determinants of child mortality using NFHS-3 data for the state of Rajasthan. Further studies using NFHS data for different states can be carried out to identify the risk factors for modelling structured dispersion present in the data.

References

1. OECD (2016) Child Mortality: http://www.oecd.org/ social/family/ database.html
2. Researchgate (2016) Infant Mortality: https://www. researchgate.net/ publication/20261350 Infant mortality.
3. Zee News (2016) Infant mortality rate- http:// zeenews.india.com/tags/ Infant-Mortality-Rate.html
4. Bharat Rakshak (2016) Indian Health Care Sector: http://forums.bharat-rakshak.com/viewtopic.php? t=5375andstart=440
5. Researchgate (2016) Socioeconomic Determinants of Infant Mortality Rate in India: https://www.researchgate.net/profile/Manak Singariya/ publication/236900863 Socioeconomic Determinants of Infant Mortality Rate in India/links/ 0deec51a2c83298b3f000000.pdf?origin=publication detail
6. Rabindranath Das etal (2011) Infant Mortality in India: Evaluating Log-Gaussian and Gamma Distributions, *The open Demography Journal*, 4, 34-41
7. Firth D. (1988) Multiplicative errors: log-normal or Gamma? *Journal of Royal Statistical Society*, 50(4), 266-268.
8. Mc Gullah P. and Nelder J.A. (1989) Generalized Linear Models, Chapman and Hall, London.
9. Myers R.H., Montgomery D.C. and Vining G.G. (2010). Generalized Linear Models with Application in Engineering and Sciences. John Wiley and Sons, New York.
10. Das (2014). Robust response surfaces regression and positive data analyses, CRC press, London, New York.
11. Grover G., Sabharwal A and Mittal J. (2013). A Bayesian approach for estimating onset time of nephropathy for type 2 diabetic patients under various health conditions. *IJSP*, 2(2), 89-101.
12. Lee Y. and Nelder J.A. (2003). Robust designs via Generalized linear models, *Journal Qual. Tech.*, 35(1), 2-12.
13. Lee Y. and Nelder J.A. (1998). Generalized linear models for the analysis of quality improvement experiments. *Can J Stat*, 26(1), 95-105.
14. Rchiips (2016) NFHS:http://rchiips.org/nfhs