M. V. Lavanya / Afr. J. Bio. Sc. 6(4) (2024) 1309-1322 ISSN: 2663-2187
https://doi.org/10.33472/AFJBS.6.4.2024.1309-1322

African Journal of Biological Sciences

A

A STUDY ON THE HURDLE POISSON REGRESSION MODEL FOR REPRODUCTIVE PATTERNS ON COUNT DATA

M. V. Lavanya[1], B. Muniswamy[2]

[1]Research Scholar, Department of Statistics, Andhra University, Visakhapatnam, Andhra Pradesh, India
[2]Professor, Head of the Department, Department of Statistics, Andhra University, Visakhapatnam, Andhra Pradesh, India. , munistats@gmail.com

Corresponding Author: *M. V. Lavanya*
*Research Scholar, Department of Statistics, Andhra University, Visakhapatnam, Andhra Pradesh, India
Email: lavanya.mv9@gmail.com

Article History
Volume 6, Issue 4, July 2024
Received: 15 jun 2024
Accepted: 25 july 2024
Published: 15 aug 2024
doi: 10.48047/AFJBS.6.4.2024.1309-1322

ABSTRACT

Children ever born (CEB) is a prevalent problem in all regions for women and has resulted in a major drop in fertility. The problem of modeling count data with lower zeros and under-dispersion, which occur in various disciplines, limits the use of typical models for count results. This study is intended to determine the impact of the CEB count on factors influenced between 2019 and 2020. This study used secondary data from the National Family Health Survey. Data was collected from 10522 women in Andhra Pradesh between the ages of 15 and 50. The Hurdle Poisson regression model was the best, revealing that women's age and fertility preference significantly influenced CEB in AP fertility in India. It is important to identify the causes of CEB to develop policies that integrate health, education, economic, and social activities to create a supportive environment for women to make informed decisions about their reproductive health.

Keyword: Children ever born, Hurdle Poisson regression, Hurdle Generalized Poisson regression, Fertility,NFHS, Under-dispersion

1. Introduction

Count data is extensively used in various academic fields, including economics, healthcare, management, industries, and other areas [13]. Mortality, migration, and fertility are the primary factors contributing to population change in any location [20]. Fertility is the most crucial factor in determining population changes [17,30]. The number of children ever born (CEB) is a crucial measure of fertility, along with the crude birth rate, total fertility rate (TFR), and age-specific fertility rate. CEB presents a complete picture of female childbirth by taking into account a woman's lifetime fertility. When it comes to influencing the growth, size, structure, and composition of the population [33] in any given region, CEB plays a significant role. The present research is designed to evaluate the occurrence of CEB and its associated factors among reproductive-age women in Andhra Pradesh. The Total Fertility Rate (TFR) in AP experienced a decrease from 2.6 in 1992-1993 to 2.3 in 1998-1999, further declining to 1.8 in 2005-2006, remaining at 1.8 in 2015-2016, and finally reaching 1.7 in 2019-2021 [35] children per women.

A common technique for modeling count data [8] in demography is Poisson regression. This technique is frequently used to evaluate variables such as the number of births, deaths, and migrations [1]. A significant problem of the Poisson model is that its underlying distributional form is limiting. According to the Poisson model, a data set with equal variance and mean is said to exhibit equi-dispersion. Data based on actual counts frequently do not correspond to this restrictive assumption; the data are either over- or under-dispersed or both. In data analysis, over-dispersion [14] can lead to underestimating standard errors, leading to overestimating the statistical significance of associated coefficients. The under-dispersion technique involves overestimating the standard errors, which underestimates the statistical significance of the explanatory variable. Over-dispersion is more common in count data; due to added uncertainty associated with using the Poisson model, practitioners frequently use the negative binomial model as an alternative [2]. Although the Poisson and negative binomial models are the primary components of count data [29,31], various extensions to these models accommodate unique aspects of the available data. The many extensions that fall into this category are zero inflation [6,18,22], zero truncation, hurdle effects, sample selection, and many others [5,25].

The count variable represents the number of CEBs in the family unit, including women aged 15 to 50, within a sample of families from Andhra Pradesh, India. Because families do not have children at random, this variable follows a Poisson distribution [3] concerning its distribution. We anticipate that many women will experience a negative contagion effect that adds to the data under dispersion: as their family size increases, their desire for additional children decreases. However, this effect will not be a general behavior across the entire population, nor will it have the same impact on all women. The impact of this effect differs based on various socio-economic and demographic factors [10,21], including the type of cooking fuel used, wealth index, place of residence, caste, religion, and more. Therefore, there are structural zeros in the data because some women may be unable to have children due to biological reasons, or they may choose not to have children for personal or socio-economic reasons. Both of these scenarios can lead to the absence of children. These zeros are not random; they result from specific circumstances that prevent childbearing. This study aims to explore these patterns of inequality and understand how various socio-economic and demographic factors influence the "reproductive choices made by women". We are interested in discovering the underlying trends and factors that influence fertility choices by investigating the preferences of individuals of varying ages and socioeconomic backgrounds about having children [7,11]. This analysis has provided insights that can be used to inform targeted interventions and policies designed to assist women's reproductive health and family planning needs.

The primary aim of this paper is to provide a model for representing datasets that contain structural zeros and may exhibit under-dispersion. We look at two statistical models: hurdle Poisson (HP), which allows for both over- and under-dispersion and hurdle generalized Poisson (HGP), which allows for under-dispersion due to the negative contagion effect.

2. Data

The fifth National Family Health Survey (NFHS-5), conducted in 2019-21, offers data on nutrition, health, and population for all states and union territories in India [24]. The NFHS [19] fourth round was held in 2015-16, five years ago. Similar to NFHS-4, NFHS-5 offers district-level estimates for numerous significant indicators. The 5th round of the NFHS provides crucial information on reproductive and child health, including socioeconomic characteristics, fertility, early child mortality, family planning, water and sanitation, nutritional status, child immunization, gender-based violence, women's empowerment, certain non-communicable diseases (NCDs), and many other topics. The Ministry of Health and Family Welfare (MoHFW), the Government of India, coordinated all five NFHS surveys [23]. The country's demographic and health database would be further strengthened by the NFHS-5 National Report, which was co-prepared by the International Institute for Population Sciences (IIPS), Mumbai, and the Statistics Division of MoHFW [15]. The Demographic and Health Surveys (DHS) Program [16], which USAID, ICF, USA fund, offered technical assistance.

The NFHS-5 fieldwork in India was conducted in two phases: Phase-I covered 17 states and 5 UTs between June 17, 2019 and January 30, 2020, and Phase-II included 11 states and 3 UTs, from January 2, 2020 to April 30, 2021 [24]. Information was gathered from 636,699 households, including 724,115 women and 101,839 men. 17 field agencies carried out the fieldwork. Four survey questionnaires—household, woman's, men, and biomarker-were used to collect information in 19 languages using Computer Assisted Personal Interviewing (CAPI). Interviews were accessible to all women aged 15-49 and males aged 15-54 [26] who resided in the selected sample households. Sigma Research and Consulting Pvt. Ltd. conducted NFHS-5 investigations in all 13 state districts of Andhra Pradesh from July 2, 2019, to November 14, 2019. The data was collected from 10,975 women [36].

3. Methodology

3.1 Regression Models

Generalized linear models (GLM) are necessary because count response variables do not have normal distributions. GLMs extend basic linear regression models to include non-normal response distributions. The random component, linear predictor, and link function are the three components of GLM, as provided by:

$$f(\mu) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \qquad (1)$$

where $X_1, X_2, \ldots, X_n$ are predictor variables, β_i, i = 0, 1, 2, ..., n are the intercept and regression coefficients. The parameter μ represents the link function.

A hurdle model comprises two main components: a point mass at zero and a distribution that produces counts greater than zero. The first component is a component that generates binary values [28], either zeros or ones. The second component generates values from a distribution that excludes zero, resulting in non-zero values. This study examines two models: Hurdle Poisson (HP) and Hurdle Generalized Poisson (HGP).

3.1.1 Hurdle Poisson (HP) Regression Model

Consider a regression model [12] in which the variable Y denotes the number of CEBs given to a woman of reproductive age in Andhra Pradesh, denoted by the variable Y. The model includes independent variables $X_1, X_2, \ldots, X_{10}$. We will use an HP regression model to analyze this data, with the response variable Y following a specific distribution.

$$\Pr(Y_i = y_i) = \begin{cases} p_i, & y_i = 0 \\ (1-p_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!\,(1-e^{-\mu_i})}, & y_i > 0 \end{cases} \qquad (2)$$

Where $0 \leq p_i < 1$, the mean and variance of this model is

$$E(Y_i) = \frac{(1-p_i)e^{-\mu_i}}{1-e^{-\mu_i}},$$

$$V(Y_i) = \frac{(1-p_i)(\mu_i + \mu_i^2)}{(1-e^{-\mu_i})} - \left[\frac{(1-p_i)\mu_i}{(1-e^{-\mu_i})}\right]^2$$

3.1.2 Hurdle Generalized Poisson (HGP) Regression Model

Let a count response $Y_i \sim HGP(\mu,\alpha)$ [27,34], i =1, 2, …, n, then Y_i has a probability function:

$$\Pr(Y_i = y_i) = \begin{cases} p_i, & y_i = 0 \\ (1-p_i)\dfrac{\left[\dfrac{\mu_i}{1+\alpha\mu_i}\right]^{y_i}\dfrac{(1+\alpha y_i)^{y_i-1}}{y_i!}\exp\left(-\dfrac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right)}{1-e^{\left(\frac{-\mu_i}{1+\alpha\mu_i}\right)}}, & y_i = 1,2,\ldots \end{cases} \quad (3)$$

In this model, the mean and variance are as follows:

$$E(Y_i) = \frac{(1-p_i)\mu_i}{1-e^{\left(\frac{-\mu_i}{1+\alpha\mu_i}\right)}},$$

$$V(Y_i) = \frac{(1-p_i)[(\mu_i(1+\alpha\mu_i)^2+\mu_i^2)]}{1-e^{\left(\frac{-\mu_i}{1+\alpha\mu_i}\right)}} - \left[\frac{(1-p_i)\mu_i}{1-e^{\left(\frac{-\mu_i}{1+\alpha\mu_i}\right)}}\right]^2.$$

3.2 Accessing Model Adequacy and Model Comparisons

The loglikelihood, Akaike Information Criterion (AIC) [4], and Bayesian Information Criterion (BIC) were then compared for all models to evaluate and select the most suitable model. The statistical tests were analyzed using the statistical software programming R 4.3.2 [9] and SPSS 29.0. The model with the minimum information criterion value was chosen as the analysis final model [9] based on the larger log-likelihood.

3.3 Variable Description and Coding

In this study, the number of CEB was used to measure the response variable of fertility, which was the outcome variable of interest. The number of children that have ever been born is a count variable that serves as an index of recent fertility. Environmental, socioeconomic, and sociodemographic factors are among the predictor variables. These variables include religion, place of residence, caste, women's age in years, place of delivery, husband age, wealth index combined, current marital status, fertility preference, and cooking fuel. In this study, the term "CEB" refers to the total number of children a woman had previously given birth to while she was still alive at the time of the survey. All women of reproductive age were included in the study because it aimed to predict women's fertility regardless of the characteristics of their backgrounds. The variables were coded according to the information provided in Table 1.

Table 1: Description of Variables

	Variables	Description
1	Type of place of residence	1 = Urban, 2 = Rural
2	Religion	1 = Hindu, 2 = Muslim, 3 = Christian
3	Type of cooking fuel	1 = Electricity, 2 = LPG, 3 = Natural gas, 4 = Biogas, 5 = Kerosene, 6 = Coal, lignite, 7 = Charcoal, 8 = Wood, 9 = Straw/Shrubs/Grass, 10 = Agricultural crop, 11= Animal dung, 96= Other
4	Wealth index combined	1 = Poorest, 2 = Poorer, 3 = Middle, 4 = Richer, 5 = Richest
5	Place of delivery	1= Home, 2 = Public, 3 = Private
6	Women age	1 = 15-19, 2 = 20-24, 3 = 25-29, 4 = 30-34, 5 = 35-39, 6 = 40-44, 7 = 45-50

7	Current marital status	1 = Single, 2 = Married, 3 = Widowed, 4 = Divorced
8	Fertility preference	1 = Have another, 2 = Undecided, 3 = No more, 4 = Sterilized, 5 = Declared infecund, 6= Never had sex
9	Caste	1 = Schedule Caste, 2 = Schedule Tribe, 3 = OBC
10	Husband age	1 = 18-27, 2 = 28-37, 3 = 38-47, 4 = 48-57, 5 = 58 & above

4. Results

Table 2 presents the descriptive statistics for the number of CEB, which is the response variable. The number of CEB ranges from 0 to 4. There were 10,522 observations, with their mean and variance of the number of CEB being 1.693 and 1.325, respectively. These values indicate that the data set shows under-dispersion. The minimum number of children was 0, while the greatest was 4. The presence of a high number of zeros (23%) is a contributing factor to the under-dispersion observed in the data set.

Table 2: Descriptive statistics of the number of CEB

Variable	N	Mean	Variance	Minimum	Maximum	Zero	Non-Zero
CEB	10522	1.693	1.325	0	4	2435(23%)	8087(77%)

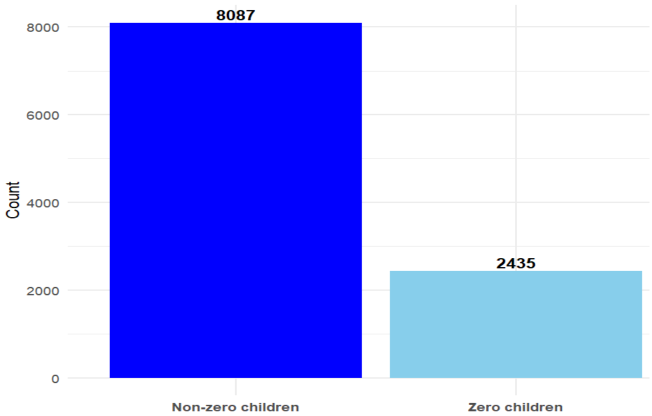

Figure 1: Zero and Non-Zero for Number of CEB

Figure 1 displays the distribution of zero and non-zero counts for the number of CEB. Specifically, there are 2435 occurrences with zero values and 8087 with non-zero values.

The number of CEB and their frequencies, along with the corresponding percentages, are presented in Table 3. It can be seen from the table that 44% of women had 2 children, making it the most common number of children. Additionally, 16% of women had 3 children, the second highest. Furthermore, it is evident that in AP, the number of women with 2-3 children exceeded those with only one child, more than 4 children, or no children.

Table 3: Frequency distribution of CEB

CEB	0	1	2	3	4 & above
Frequency	2435	1200	4627	1682	578
Percent	23.1	11.4	44.0	16.0	5.5

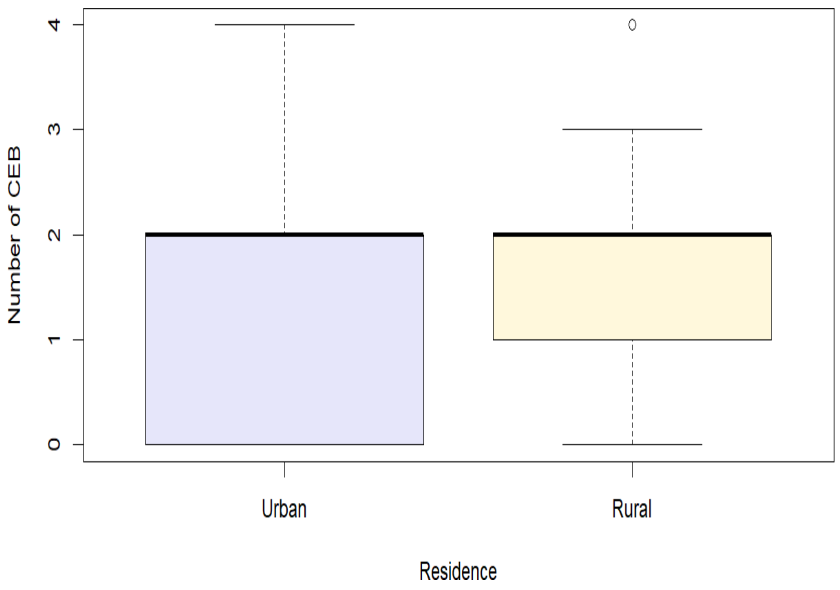

Figure 2: Boxplot for the Number of CEB by Residence

Figure 2 displays a boxplot illustrating the number of CEB based on the location of women, namely urban and rural areas. The data belongs to women who have given birth to two children.

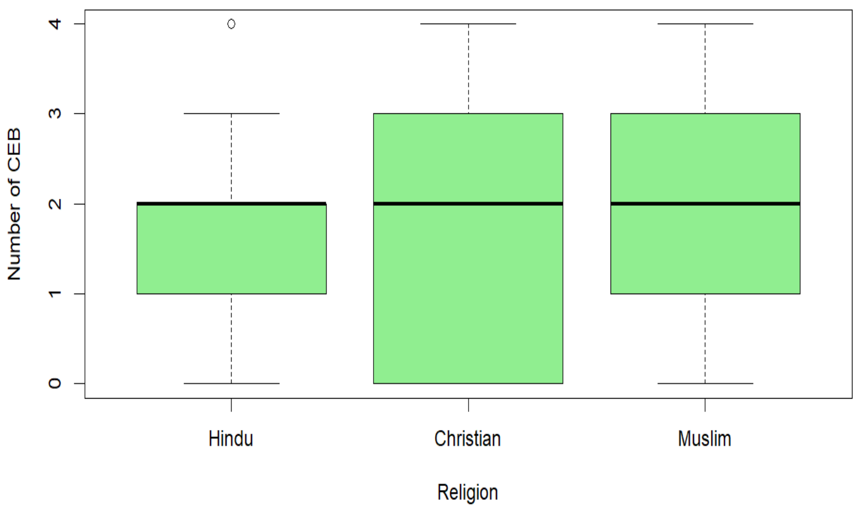

Figure 3: Boxplot for Number of CEB by Religion

Figure 3 presents a boxplot illustrating the distribution of CEB among women of different religions, namely Hindu, Christian, and Muslim. The data includes women who have given birth to two children. Based on the comparison from table 4, it is evident that the log-likelihood of the HP model (-11640) is higher than that of the HGP model (-13983.45). This study indicates that the HP model better fits the data. The AIC of the HP model (23446.01) is lower than that of the HGP model (28128.9), suggesting that the HP model is more suitable for fit and complexity. The BIC value for the HP model (24026.91) is lower than that of the HGP model (28717.06), suggesting that the HP model performs better in terms of BIC. BIC is known for penalizing model complexity more rigorously than AIC. Therefore, compared to the HGP model, the HP model performs more on these criteria.

Table 4: Overall model comparison by model fit characteristics

Test Statistics	HP	HGP
Log Likelihood	-11640	-13983.45
AIC	23446.01	28128.9
BIC	24026.91	28717.06

The outcomes of the modeling of the number of CEB using HP and HGP for count models are shown in Table 5. The reference group in this instance consists of unmarried Hindu women, ages 15 to 19, who belong to a scheduled caste, use electrical fuel, live in urban areas with the lowest levels of affluence, consider having another child with a partner who is between the ages of 18 and 27, and have given birth at home during the year before to the interview. The following predictors were statistically significant for the regression model part that predicted the number of CEB: women's age, fertility preference (excluding never having sex), and the Muslim religion, which has a positive, richer, and richest wealth index that is negative.

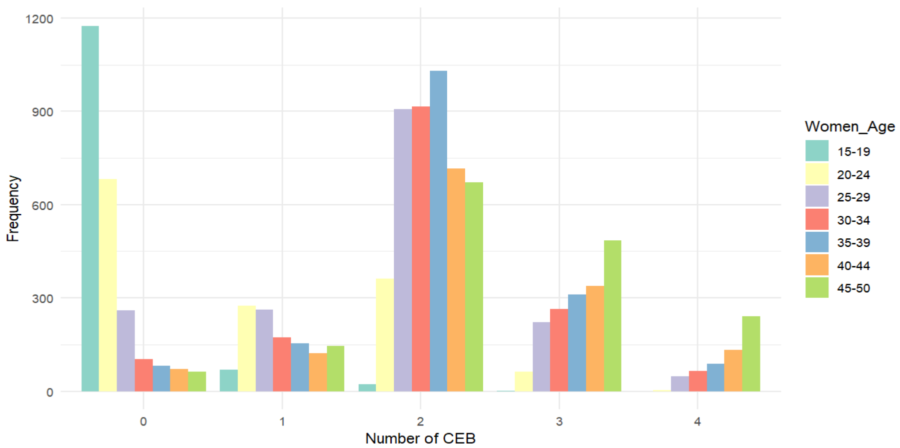

Figure 4: Cross table plot for the Number of CEB with Women's age

Figure 4 illustrates the cross-tabulation of the age of women and the number of CEBs. The data indicates that the women had two children between the ages of 35-39, followed by children born, while the women were between the ages of 25-29 and 30-34.

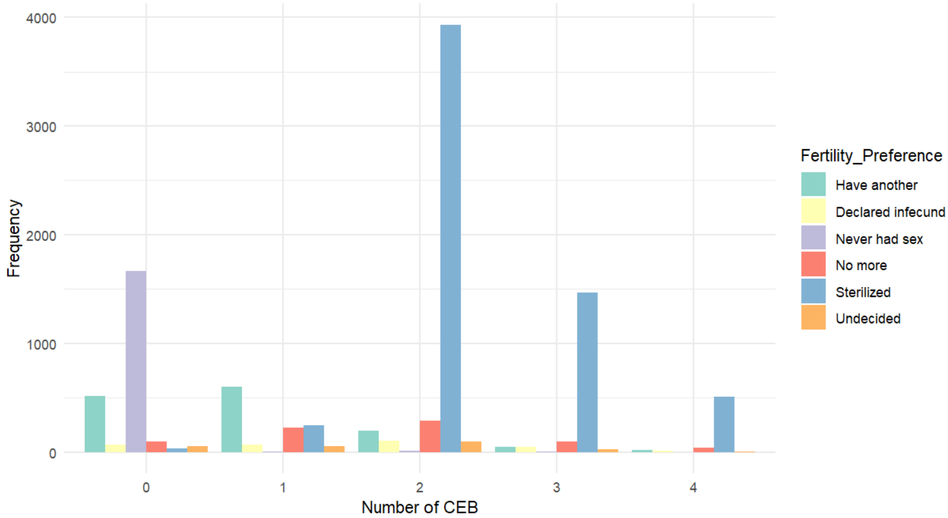

Figure 5: Cross table plot for Number of CEB with Fertility Preference

Figure 5 depicts the cross-tabulation of fertility choice and the number of CEB. The data suggests that women who had two children had sterilization due to their choice to limit reproduction.

In the HP and HGP models, women residing in rural regions have a lower (IRR = 1.003, 95% CI: 0.958–1.050) and (IRR = 1.00, 95% CI: 0.963–1.038) count of CEB in comparison to those living in urban areas. Muslim women are substantially more likely than Hindu women to have a higher count of CEB in 16% (IRR = 1.61, 95% CI: 1.087–1.241) and 10% (IRR = 1.10, 95% CI: 1.041–1.162) of both models. Furthermore, Christians show a small increase of 6.6% (IRR = 1.066, 95% CI: 0.998–1.139) compared to Hindus only in the HP model.

Compared to women who use electricity fuel, those who use biogas cooking fuel had lower CEB counts in HP (IRR = 0.539, 95% CI: 0.210–1.382) and HGP (IRR = 0.730, 95% CI: 0.394–1.352). With 17% (IRR = 0.835, 95% CI: 0.756–0.923) or 27% (IRR = 0.727, 95% CI: 0.650–0.814) and 11% (IRR = 0.890, 95% CI: 0.819–0.967) or 18% (IRR = 0.82, 95% CI: 0.749–0.898) having a lower count of CEB compared to the poorest, HP and HGP are in the richer or richest wealth index groups. In addition, women who gave birth in public showed a decline in both models, while those who gave birth in the private health sector showed an increase in the number of CEB in HP compared to women who gave birth at home.

Women aged 20–24 to 45–50 had greater CEB counts than those from 15–19 years old. When comparing the IRR of a count of CEB to women aged 15–19, the results were as follows: 2.265 (IRR = 3.265, 95% CI: 2.689–3.966), 2.056 (IRR = 3.056, 95% CI: 2.515–3.714), 1.855 (IRR = 2.855, 95% CI: 2.351–3.466), 1.790 (IRR = 2.790, 95% CI: 2.297–3.389), 1.763 (IRR = 2.763, 95% CI: 2.277–3.352), and 1.386 (IRR = 2.386, 95% CI: 1.963–2.900) times higher among 45–50, 40–44, 35–39, 30–34, 25–29, and 20–24, respectively. The results show that marital status significantly predicts the count of CEB in HGP; the count of CEB for married women is 3.113 times (IRR = 4.113, 95% CI: 3.083–5.487), which is higher than for single women.

It was also seen that the incidence rate ratio (IRR) of the number of CEB went up a lot among undecided women, with no more children, were sterilized, or declared infecund, compared to women who wanted another child in both models. On the other hand, women who reported never having had sex showed a significantly lower count of CEB in the HGP model. Women belonging to the Scheduled Tribe and Other Backward Classes had a lower count of CEB compared to women belonging to the Scheduled Caste in both models. The reproductive rate of women, dependent on the age of their husbands, is 1.7% (IRR = 1.017, 95% CI: 0.961–1.076) for ages 38–47 in the HP group and 2.6% (IRR = 1.026, 95% CI: 0.980–1.074) for ages 48–57 in the HGP group. These rates are greater than those of husbands aged 18–27, resulting in a larger count of CEB.

Table 5: Results from HP and HGP Model: Count Model Coefficients

Variables	Category	HP				HGP			
		P-value	IRR	95% Wald Confidence Interval I RR		P-value	IRR	95% Wald Confidence Interval IRR	
Intercept1		0.000***	0.337	0.198	0.573	0.000***	0.095	0.065	0.139
Intercept2		–	–	–	–	0.997@	0.000	0.000	Inf
Place of residence (Ref: Urban)	**Rural**	0.907@	1.003	0.958	1.050	0.994@	1.000	0.963	1.038

Religion (Ref: Hindu)	**Muslim**	0.000***	1.161	1.087	1.241	0.001***	1.100	1.041	1.162
	Christian	0.059*	1.066	0.998	1.139	0.264@	1.032	0.977	1.090
Type of cooking fuel (Ref: Electricity)	**LPG**	0.716@	0.961	0.778	1.188	0.969@	0.997	0.838	1.186
	Biogas	0.198@	0.539	0.210	1.382	0.317@	0.730	0.394	1.352
	Kerosene	0.513@	1.184	0.714	1.963	0.808@	1.054	0.691	1.607
	Coal, lignite	0.578@	0.875	0.545	1.402	0.887@	0.973	0.666	1.420
	Charcoal	0.714@	1.049	0.811	1.359	0.522@	1.072	0.866	1.329
	Wood	0.992@	0.999	0.804	1.241	0.701@	1.036	0.866	1.238
	Straw/shrubs/grass	0.804@	1.043	0.749	1.452	0.725@	1.051	0.797	1.385
	Agricultural crop	0.811@	0.964	0.716	1.298	0.941@	0.991	0.777	1.264
	Animal dung	0.831@	0.869	0.239	3.159	0.925@	0.954	0.352	2.583
	Other	0.585@	0.698	0.192	2.539	0.677@	0.809	0.299	2.193
Wealth index combined (Ref: Poorest)	**Poorer**	0.140@	0.931	0.848	1.024	0.342@	0.963	0.890	1.041
	Middle	0.014**	0.887	0.806	0.976	0.106@	0.936	0.865	1.014
	Richer	0.000***	0.835	0.756	0.923	0.006***	0.890	0.819	0.967
	Richest	0.000***	0.727	0.650	0.814	0.000***	0.820	0.749	0.898
Place of delivery (Ref: Home)	**Public**	0.453@	0.986	0.949	1.024	0.424@	0.988	0.958	1.018
	Private	0.943@	1.003	0.934	1.077	0.683@	0.988	0.932	1.047
Women age (Ref: 15-19)	**20-24**	0.003***	1.827	1.231	2.711	0.000***	2.386	1.963	2.900
	25-29	0.000***	2.128	1.440	3.146	0.000***	2.763	2.277	3.352
	30-34	0.000***	2.173	1.469	3.214	0.000***	2.790	2.297	3.389
	35-39	0.000***	2.246	1.519	3.321	0.000***	2.855	2.351	3.466
	40-44	0.000***	2.472	1.671	3.657	0.000***	3.056	2.515	3.714
	45-50	0.000***	2.721	1.840	4.023	0.000***	3.265	2.689	3.966
Current marital status (Ref: Single)	**Married**	0.126@	1.240	0.942	1.633	0.000***	4.113	3.083	5.487
	Widowed	0.323@	1.152	0.870	1.527	0.000***	3.799	2.836	5.089
	Divorced	0.276@	0.821	0.559	1.181	0.000***	2.378	1.694	3.338
Fertility preference (Ref: Have another)	**Undecided**	0.000***	1.868	1.564	2.230	0.000***	1.585	1.403	1.792
	No more	0.000***	1.923	1.691	2.186	0.000***	1.744	1.607	1.894
	Sterilized	0.000***	2.544	2.282	2.837	0.000***	2.252	2.111	2.402
	Declared infecund	0.000***	1.838	1.562	2.163	0.000***	1.489	1.331	1.667
	Never had sex	0.000***	2.144	1.463	3.141	0.000***	0.175	0.118	0.260
Caste (Ref: Schedule caste)	**Schedule tribe**	0.232@	0.959	0.896	1.027	0.429@	0.978	0.925	1.034
	OBC	0.160@	0.970	0.930	1.012	0.485@	0.988	0.954	1.022
Husband age (Ref: 18-27)	**28-37**	0.476@	1.021	0.964	1.082	0.534@	1.015	0.968	1.064
	38-47	0.565@	1.017	0.961	1.076	0.354@	1.022	0.976	1.070
	48-57	0.522@	1.019	0.963	1.078	0.267@	1.026	0.980	1.074
	58 & above	0.435@	1.023	0.966	1.084	0.430@	1.019	0.972	1.068

***1% Level of Significant (p-value<0.01)

** 5% Level of Significant (p-value<0.05)

* 10% Level of Significant (p-value<0.1)

@ Not Significant

The data in Table 6 indicates the odds of women not having children, as determined by the HP and HGP models. The variables of place of residence, religion, type of cooking fuel, wealth index, caste, and husband's age do not significantly impact the odds of having no children. The odds of individuals aged 25-29 and 30-34 having no children are 92.9 and 92.6 times greater than those in the 15-19 age group. The likelihood of having no children is much higher among older age groups (20-50). The probability of married women never having any children is predicted to be 5.04 times greater than that of unmarried women. Having a preference for sterilization, no more children, or being undecided was associated with 5.2, 1.77, and 1.1 times higher chances of not having any children, respectively, compared to having a preference for having another child. When compared to giving birth at home, giving birth in private facilities lowers the likelihood of having no children. This study suggests that access to healthcare facilities, as well as preferences for institutional childbirth, may have an impact on fertility rates and family planning beliefs.

Table 6: Results from HP and HGP Model: Zero Hurdle Model Coefficients

Variables	Category	Estimate	Std. Error	z-value	P-value	IRR	95% Wald Confidence Interval IRR	
Intercept		−6.529	0.649	−10.07	0.000***	0.001	0.000	0.005
Place of residence (Ref: Urban)	Rural	0.038	0.120	0.319	0.749@	1.039	0.821	1.315
Religion (Ref: Hindu)	Muslim	−0.068	0.186	−0.363	0.717@	0.935	0.649	1.347
	Christian	−0.287	0.160	−1.790	0.074*	0.750	0.548	1.028
Type of cooking fuel (Ref: Electricity)	LPG	0.228	0.500	0.456	0.648@	1.256	0.471	3.347
	Biogas	0.334	1.361	0.246	0.806@	1.397	0.097	20.141
	Kerosene	−0.371	1.029	−0.361	0.718@	0.690	0.092	5.184
	Coal, lignite	2.596	2.833	0.916	0.360@	13.414	0.052	3461.45
	Charcoal	0.519	0.668	0.777	0.437@	1.680	0.454	6.215
	Wood	0.555	0.519	1.069	0.285@	1.742	0.630	4.814
	Straw/shrubs/grass	−0.020	0.905	−0.023	0.962@	0.980	0.166	5.769
	Agricultural crop	0.096	0.699	0.138	0.891@	1.101	0.280	4.334
	Animal dung	9.897	577.61	0.017	0.956@	19873.66	0.000	Inf
	Other	7.775	622.17	0.012	0.996@	2379.74	0.000	Inf
Wealth index combined (Ref: Poorest)	Poorer	0.233	0.240	0.971	0.331@	1.262	0.789	2.019
	Middle	0.390	0.246	1.594	0.111@	1.477	0.941	2.386
	Richer	0.148	0.256	0.578	0.564@	1.159	0.702	1.913
	Richest	0.300	0.278	1.077	0.281@	1.350	0.782	2.328
Place of delivery (Ref: Home)	Public	−0.094	0.100	−0.943	0.346@	0.910	0.748	1.107
	Private	−0.346	0.175	−1.974	0.048*	0.707	0.502	0.998
Women age (Ref: 15-19)	20-24	1.311	0.168	7.818	0.000***	3.711	2.671	5.156
	25-29	1.929	0.179	10.769	0.000***	6.686	4.847	9.783
	30-34	1.926	0.209	9.222	0.000***	6.862	4.557	10.333
	35-39	1.862	0.212	8.719	0.000***	6.435	4.249	9.746
	40-44	1.812	0.232	7.812	0.000***	6.120	3.885	9.642
	45-50	1.779	0.236	7.554	0.000***	5.926	3.735	9.404
Current marital status (Ref: Single)	Married	5.038	0.306	16.457	0.000***	154.199	84.623	280.98
	Widowed	4.084	0.352	11.587	0.000***	59.377	29.758	118.47
	Divorced	2.719	0.392	6.934	0.000***	15.160	7.030	32.694
Fertility preference (Ref: Have another)	Undecided	1.172	0.217	5.398	0.000***	3.230	2.110	4.943
	No more	1.777	0.161	11.048	0.000***	5.912	4.313	8.103
	Sterilized	5.201	0.253	20.527	0.000***	181.425	110.4	298.10
	Declared infecund	0.876	0.197	4.440	0.000***	2.401	1.631	3.534
	Never had sex	−0.284	0.355	−0.801	0.423@	0.753	0.375	1.509
Caste (Ref: Schedule caste)	Schedule tribe	0.110	0.172	0.643	0.520@	1.119	0.797	1.564
	OBC	0.129	0.111	1.164	0.244@	1.138	0.916	1.414
Husband age	28-37	0.013	0.152	0.087	0.931@	1.013	0.752	1.366

(Ref: 18-27)	**38-47**	0.243	0.148	1.639	0.101@	1.275	0.954	1.706
	48-57	0.219	0.148	1.480	0.139@	1.245	0.931	1.664
	58 & above	0.054	0.151	0.358	0.721@	1.056	0.785	1.419

***1% Level of Significant (p-value<0.01)

** 5% Level of Significant (p-value<0.05)

* 10% Level of Significant (p-value<0.1)

@ Not Significant

5. Discussion

The study used the HP and HGP regression models to analyze fertility patterns among women aged 15-50 years in Andhra Pradesh; based on the CEB data from NFHS-5, several factors were considered in this analysis. The model involves various kinds of socio-demographic, socio-economic, and environmental variables. This study involved 10,522 women, with 2,435 (23%) having no children and the remaining 8,087 (77%) having at least one child. The data set had a small number of zeros, which led to the under-dispersion.

According to this study, there is a higher frequency of women who have given birth to two children. Variations in healthcare access, education, employment prospects, and cultural attitudes towards family planning and size can explain the urban-rural disparity in women's fertility rates. The CEB counts of married women were substantially greater than those of women with reproductive choices, such as sterilization or no more children. Women from Scheduled Tribes and Other Backward Classes had a lower count of CEB compared to those from Scheduled Castes. This study shows that caste-based socioeconomic discrepancies may impact reproduction choices, presumably due to differences in access to resources, education, and healthcare. The husband's age also had an impact, as older husbands were found to have higher CEB counts.

This study also found that several factors did not significantly impact the predictions. These factors included place of residence, religion, type of cooking fuel, wealth index, caste, and husband's age. However, age and marital status were shown to be important factors. Women aged 25-34 had considerably greater probabilities of having no children than those aged 15-19. Married women were more likely to have no children than unmarried women. Preferences for sterilization, no more children, or remaining undecided were all connected with a greater likelihood of not having children. This result shows that how a family chooses to use energy can have health and environmental effects on how many children they have. It also suggests that there may be connections between health knowledge, financial status, and decisions about having children. Changes in reproductive choices and household decision-making processes, which are impacted by spouse age and generational transitions, may cause this gap. Higher wealth index categories have lower fertility rates, which indicates that economic stability, availability of resources, and lifestyle choices all play a role in the decision-making process regarding reproduction.

6. Conclusion

This study aimed to determine significant observations on the demographic and socioeconomic variables that impact the count of CEB among women. This study was achieved through count data regression analysis, employing data from the National Family Health Survey (NFHS) conducted between 2019 and 2021. The HP regression model was identified as the most optimal model for the dataset and displayed under-dispersion. It demonstrated that women's age, fertility preference, and Muslim faith were major factors in determining the count of the number of CEB in AP. Age, marital status, and fertility choice of women are crucial determinants of infertility. Women living in rural areas who use a specific type of cooking fuel, receive deliveries at public institutions, and belong to a particular caste have a reduced probability of having a non-zero number of CEB.

Therefore, women must have a role in enabling rural areas to receive electricity, ensuring safe deliveries at healthcare facilities, and implementing programs that promote women's education. The government should have complete fertility policies that consider all the different things that affect people's decisions about having children. These policies should incorporate health, education, economic, and social efforts to establish a supportive environment for women to make educated choices regarding their reproductive health.

Ethics approval and consent to participate

The authors can download survey data from the Demographic and Health Surveys (DHS) Programme. The data is publicly available and does not contain any personal information.

Acknowledgments

The author's appreciation goes to the DHS Program that granted access to use the data analyzed in the study.

References

1. Abdulkabir, M., Edem, U., Tunde, R., & Kemi, B. (2015). An empirical study of generalized linear model for count data. *Journal of Applied and Computational Mathematics*, *4*, 253.
2. Agresti, A. (2019). An introduction to categorical data analysis.
3. Al-Balushi, M. S., Ahmed, M. S., Islam, M. M., & Khan, M. H. R. (2020). Multilevel poisson regression modeling to identify factors influencing the number of children ever born to married women in Oman. Journal of Statistics and Management Systems, 23(8), 1357-1373.
4. Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle-In Second International Symposium on Information Theory (Eds) BN Petrov, F. Csaki. BNPBF Csaki Budapest: Academiai Kiado.
5. Aragaw Eshetie Aguade and B. Muniswamy. 2018. Proposed Score Test for Over-dispersion Parameter in the Multilevel Negative Binomial Regression Model, Journal of Emerging Technologies and Innovative Research, Volume 5, Issue 12, pp. 709-720, ISSN-2349-5162.
6. Aye, H. H. (2023). Analysis of Under-Five Mortality of Myanmar Using Zero-Inflated Regression Models (Hla Hla Aye, 2023) (Doctoral dissertation, MERAL Portal).
7. Bongaarts, J. 1 978. A framework for analyzing the proximate determinants of fertility. Population and Development Revie1v, 4(1), 105-132.
8. Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (No. 53). Cambridge university press.
9. Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and recent advances in model selection. Philosophy of statistics, 583-605.
10. Das, P., Das, T., & Roy, T. B. (2023). Role of proximate and non-proximate determinants in children ever born among Indian women: Change detection analysis from NFHS-3 & 5. Women and Children Nursing, 1(1), 9-17.
11. Finlay, J. E., Mejia-Guevara, I., & Akachi, Y. (2018). Inequality in total fertility rates and the proximate determinants of fertility in 21 sub-Saharan African countries. PloS one, 13(9), e0203344.

12. Gannoun, A., Saracco, J., & Yu, K. (2007). Comparison of kernel estimators of conditional distribution function and quantile regression under censoring. Statistical Modelling, 7(4), 329–344.
13. Greene, W. (2007). Functional form and heterogeneity in models for count data. Foundations and Trends® in Econometrics, 1(2), 113–218.
14. Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. Journal of Business & Economic Statistics, 9(2), 215–222.
15. IIPS, O. (2007). National Family Health Survey (NFHS-3), 2005-06: India. Vol. I. Mumbai: International Institute for Population Sciences.
16. India: DHS, 2019-2021 - Final Report (English) https://dhsprogram.com/publications/publication-FR375-DHS-Final-Reports.cfm.
17. Laelago, T., Habtu, Y., & Yohannes, S. (2019). Proximate determinants of fertility in Ethiopia; an application of revised Bongaarts model. Reproductive health, 16, 1-9.
18. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34(1), 1-14.
19. Iips, I. C. F. (2021). National Family Health Survey (NFHS-5): 2019-21 India. Mumbai: International Institute for Population Sciences (IIPS).
20. Mekonnen, W., & Worku, A. (2011). Determinants of fertility in rural Ethiopia: the case of Butajira Demographic Surveillance System (DSS). BMC public health, 11, 1-6.
21. Melese, Z. Y., & Zeleke, L. B. (2020). Factors affecting children ever born among reproductive aged women in Ethiopia; data from Edhs 2016. World, 5(3), 66-75.
22. Muniswamy, B., Molla, D. T., & Reddy, N. K. (2015). Comparison of test statistic for zero-inflated negative binomial against zero-inflated Poisson model. Indian Journal of Science and Technology, 8(4), 349.
23. National Family Health Survey (NFHS-4), India, 2015-16: Chhattisgarh - India Environment Portal | News, reports, documents, blogs, data, analysis on environment & development | India, South Asia. http://www.indiaenvironmentportal.org.in/content/451850/national-family-health-survey-nfhs-4-india-2015-16-chhattisgarh/
24. Raina, S. K., & Kumar, R. (2023). Need to involve relevant stakeholders-lessons from NFHS data. Journal of Family Medicine and Primary Care, 12(4), 603-605.
25. Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of biopharmaceutical statistics, 16(4), 463-481.
26. Rutstein, S. O., & Rojas, G. (2006). Guide to DHS statistics. Calverton, MD: ORC Macro, 38, 78.
27. Saffari, S. E., Adnan, R., & Greene, W. (2013). Investigating the impact of excess zeros on hurdle-generalized Poisson regression model with right censored count data. Statistica neerlandica, 67(1), 67-80.
28. Suriya, V., & Geetha, R. (2023). PREDICTION OF THE DETERMINANTS OF THE NUMBER OF ANTENATAL CARE VISITS IN NFHS IV SURVEY OF INDIA: MODELING EXCESS ZERO OF COUNT DATA. JP Journal of Biostatistics, 23(2), 173-200.
29. Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. Journal of econometrics, 84(1), 129-154.
30. Tomal, J. H., Khan, J. R., & Wahed, A. S. (2022). Weighted Bayesian Poisson regression for the number of children ever born per woman in Bangladesh. Journal of Statistical Theory and Applications, 21(3), 79-105.

31. Winkelmann, R., & Zimmermann, K. F. (1994). Count data models for demographic data. Mathematical Population Studies, 4(3), 205–221.
32. Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. Journal of statistical software, 27(8), 1–25.
33. Zelalem, D., Semahegn, A., Tesfaye, G., & Sileshi, B. (2015). The level and patterns of fertility among women in Kersa demographic surveillance and health research center (KDS-HRC) field site, Kersa District, East Ethiopia. Fertility Research and Practice, 1, 1–8.
34. Zuo, G., Fu, K., Dai, X., & Zhang, L. (2021). Generalized Poisson hurdle model for count data and its application in ear disease. Entropy, 23(9), 1206.
35. https://www.indiabudget.gov.in/economicsurvey/doc/stat/tab818.pdf
36. https://rchiips.org/nfhs/NFHS5_FCTS/COMPENDIUM/Andhra_Pradesh.pdf National Family Health Survey. NFHS-5: Compendium of Fact Sheet, KEY INDICATORS State and Districts of Andhra Pradesh. (Accessed August 24, 2023). (www.rchiips.org)