International Journal of Statistics and Applications 2023, 13(1): 1-12
DOI: 10.5923/j.statistics.20231301.01

Bayesian Method for Difference in Differences Estimation of Impact in Experimental Interventions with Count Outcome

Oladipupo B. Ipadeola[1,*], Kamil D. Ajagbe[2], Adetunji K. Ilori[3], Timothy A. Attah[4], Abiodun F. Ipadeola[5], Emmanuel T. Jolayemi[1]

[1]University of Ilorin, Ilorin, Nigeria
[2]Nasarawa State University, Keffi, Nasarawa State, Nigeria
[3]National Mathematical Center, Abuja, Nigeria
[4]National Space Research and Development Agency, Abuja, Nigeria
[5]Covenant University, Otta, Ogun State, Nigeria

Abstract One of the widely used approach for estimation of impact in experimental intervention is the Difference in Differences (DID) Regression which depends on assumptions that are often violated in real life. This study provides a Bayesian DID estimation alternative to the classical DID regression approach when the outcome variable is count. The performance of the two approaches was examined using simulated and real-life data. For the proposed Bayesian DID estimation approach, the distribution of the difference of differences of outcome among out those exposed to intervention (treatment group) and those not exposed (control group) before and after the intervention, was derived using the convolution process. The resulting distribution was a 4 parameter Skellam distribution. The likelihood of the Skellam distribution was combined with Gamma and Power priors independently to determine the posteriors. All the posterior distributions derived were intractable and parameter estimation was carried out using the Metropolis Hasting Algorithm. The impact estimates obtained from Bayesian DID estimation method using simulated data were closer to the true value and had lower Mean Squared Error than those obtained from classical DID regression approach. Result from Bayesian DID estimation approach with Power prior provided a realistic impact estimate with the lowest Mean Squared Error compared to other methods. This study revealed that impact estimated using classical DID regression overstate the reality.

Keywords Bayesian, Difference in Differences, Impact, Estimation, Prior, Posterior, Intervention

1. Introduction

In experimental interventions, decisions are made from analysis of panel data derived by collecting data from two or more time periods, before and within/after an intervention, from a group of individuals who were exposed to the intervention compared to a similar group of individuals that were not exposed to the intervention. Experimental intervention provides the approach for determining the "Average treatment effect" and the "Average effect of treatment on the treated". The average treatment effect is the average impact of the program across all the subjects in the population of interest [1]. Estimation of the average treatment effect otherwise known as impact is often conducted using classical approaches. One of the widely adopted approach is the Difference in Difference Regression.

* Corresponding author:
ladiipadeola@gmail.com (Oladipupo B. Ipadeola)
Received: Mar. 15, 2023; Accepted: Apr. 3, 2023; Published: May 12, 2023
Published online at http://journal.sapub.org/statistics

Bayesian estimation is similar to maximum likelihood in the ability of these two methods to estimate random and fixed variables [2]. However, it differs from maximum likelihood in that the posterior probability is maximized rather than likelihood function. This is given by the function of the likelihood multiplied by the prior distribution of the posterior probability. In this case, Bayesian estimate addresses the complexities of maximum likelihood. Bayesian estimate bypasses the need for the design of likelihood function and have computational tools that can essentially be used for simple statistical techniques. Bayesian largely depends on all useful information and utilizes prior information in its estimations. Therefore, this method is most ideal when prior information is available compared to maximum likelihood which ignores prior information in its parameters.

Bayesian method is not commonly used in drawing inferences in experimental interventions due to two major reasons. First, except for the simplest applications, Bayesian analyses are computationally difficult and were infeasible until fast computers and simulation-based methods were developed. Second, it can be difficult to settle on a prior that is widely accepted [3]. However, the advantages of Bayesian approach over classical approach are well documented in literatures [3-5]. A previous study has presented the procedure for estimating impact using the Bayesian Difference in Difference estimation alternative to the Bayesian regression and classical regression methods when the outcome variable is continuous and from normal distribution [6]. This paper describes the procedure for estimating impact using the Bayesian Difference in Difference estimation alternative to the Bayesian regression and classical regression methods when the outcome of interest is count and from Poisson distribution.

2. Material and Method

Suppose events are occurring randomly and uniformly in time. The events occur with a known average. Let Y be the number of events occurring in a fixed period. Then, Y will have a Poisson distribution with parameter θ such that,

$$P(y) = \frac{\theta^{y} e^{-\theta}}{Y!}, \quad y = 0,1,2,3,... \tag{1}$$

The intensity parameter, θ, represents the expected number of occurrences in a fixed period i.e. $\theta = E[Y]$. It is also the variance of the count: $\theta = Var[Y] \Rightarrow \theta > 0$.

2.1. Classical Difference in Differences Regression Approach

Let y_{ijt} be the count of events and follows Poisson distribution with parameter θ. If we have two independent groups, treatment and control, before program implementation, denote count of events in the control group and the treatment group as y_{0j0} and y_{1j0}, with means $\bar{y}_{0j0} = \bar{y}_{1j0}$ respectively, and at the end of program implementation, we denote the count of events as y_{0j1} and y_{1j1}, with means $\bar{y}_{0j1} \leq \bar{y}_{1j1}$ respectively.

If:

$$P = \begin{cases} 1 & treatment\ group \\ 0 & control\ group \end{cases}$$

and

$$T = \begin{cases} 1 & Follow\ up \\ 0 & Baseline \end{cases}$$

$$y_{ijt} \sim Pois(\theta)$$

From the generalized linear model

$$logY_{ijt} = W\theta; \forall P, T \in W$$

To determine the impact of the intervention, the model to estimate is:

$$\log Y_{ijt} = \theta_0 + \theta_1 P_{it} + \theta_2 T_{ij} + \theta_3 P_{it} T_{ij} + \zeta_{ijt} \tag{2}$$

Where, P is the Program status, T is time, $\theta_0, \theta_1, \theta_2$ and θ_3 are model parameters and θ_3 is the coefficient of the interaction between Program status and Time, which is the impact parameter.

In treatment group

$$P_i = 1$$
$$T_j = 0$$

- At Baseline:

$$\log Y_{1j0} = \theta_0 + \theta_1 + 0 + 0 + \zeta_{1j0} \tag{3}$$

$$T_j = 1$$

- At Follow-up:

$$\log Y_{1j1} = \theta_0 + \theta_1 + \theta_2 + \theta_3 + \zeta_{1j1} \tag{4}$$

Taking the difference (4) and (3):

$$\left(\log Y_{1j1} - \log Y_{1j0}\right) = \theta_2 + \theta_3 + \left(\zeta_{1j1} - \zeta_{1j0}\right) \tag{5}$$

$$E\left(\log Y_{1j1} - \log Y_{1j0}\right) = \theta_2 + \theta_3 \tag{6}$$

In control group

$$P_i = 0$$
$$P_i = 0$$

- At Baseline:

$$\log Y_{0j0} = \theta_0 + 0 + 0 + 0 + \zeta_{0j0} \tag{7}$$

$$T_j = 1$$

- At Follow-up:

$$\log Y_{0j0} = \theta_0 + \theta_2 + 0 + \zeta_{0j1} \tag{8}$$

Taking the difference:

$$\left(\log Y_{0j1} - \log Y_{0j0}\right) = \theta_2 + \left(\zeta_{0j1} - \zeta_{0j0}\right) \tag{9}$$

$$E\left(\log Y_{0j1} - \log Y_{0j0}\right) = \theta_2 \tag{10}$$

Now, the difference in differences is given by

$$E\left(\log Y_{1j1} - \log Y_{1j0} | P = 1\right)$$
$$-E\left(\log Y_{0j1} - \log Y_{0j0} \,| P = 0\right) = (\theta_2 + \theta_3) - \theta_2 \tag{11}$$

$$E\left(\log Y_{1j1} - \log Y_{1j0} | P = 1\right)$$
$$-E\left(\log Y_{0j1} - \log Y_{0j0} \,| P = 0\right) = (\theta_2 + \theta_3) - \theta_2 = \theta_3 \tag{12}$$

$$E\left(\log Y_{1j1} - \log Y_{1j0} | P = 1\right)$$
$$-E\left(\log Y_{0j1} - \log Y_{0j0} \,| P = 0\right) = \theta_3 \tag{13}$$

Therefore,

$$\theta_3 = E\left\{\log\left(\frac{\left((Y_{1j1} - Y_{1j0})\middle| P=1\right)}{\left((Y_{0j1} - Y_{0j0})\middle| P=0\right)}\right)\right\} \tag{14}$$

$$E\left(\zeta_{1j1} - \zeta_{1j0} | P = 1\right) = 0$$

And

$$E\left(\zeta_{0j1} - \zeta_{0j0} | P = 0\right) = 0$$

The percentage transformation of the impact is obtained by

$$\exp\theta_3 - 1 \tag{15}$$

Thus, the difference in difference estimator θ_3 directly approximates the causal treatment effect where $\exp\theta_3 - 1$ is the transformation from log points to percentage points [7].

2.2. Bayesian Difference in Differences Regression Approach

The Bayesian equation provides a probability distribution of θ given observations of the data yy. In this equation, $p(y)$ is the sum (or integral) of $p(\theta)p(y|\theta)$ over all possible values of θ.

Therefore,

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \quad (16)$$

Where:

$p(\theta|y)$ denote the posterior distribution of the parameter $p(y|\theta) = L(\theta)$ is the likelihood function $p(\theta)$ is the prior distribution of which express our beliefs about the parameters, before observing the data

$p(y)$ is called the marginal likelihood and plays the role of the normalizing constant of the density of the posterior distribution.

For

$$Y_{ijt} \sim Pois(\theta)$$

If there exist a conjugate prior

$$\theta \sim G(a,b)$$

Given that the likelihood of y is

$$\ell(y) = \frac{\theta^{\sum y} e^{-n\theta}}{\prod y!} \quad (17)$$

If the prior distribution is given as

$$\pi(y) = \frac{\theta^{a-1}}{b^a \Gamma(a)} e^{-\frac{\theta}{b}} \quad (18)$$

Then the posterior distribution is derived as follows:

$$\pi(\theta|y) = \frac{\theta^{\sum y} e^{-n\theta}}{\prod y!} \times \frac{\theta^{a-1}}{b^a \Gamma(a)} e^{-\frac{\theta}{b}} \quad (19)$$

$$\pi(\theta|y) = \frac{\theta^{a+\sum y-1} e^{-\frac{\theta(nb+1)}{b}}}{b^a \Gamma(a) \prod y!} \quad (20)$$

$$\theta|y = G\left(a + \sum y, \frac{b}{nb+1}\right) \quad (21)$$

Therefore,

$$a^* = a + \sum y \quad (22)$$
$$= a + n\bar{y}$$
$$b^* = \frac{b}{nb+1} \quad (23)$$

2.2.1. Choice of Prior

In many observational and experimental studies with count response data, choice of prior is often subjective [8,9]. However, use of variants of normal distribution priors for Poisson response variable is very popular and widely adopted. [10,11]. This study explored power, gamma priors and Uniform prior to estimate average treatment effect in the Bayesian context.

Using Power prior, suppose the information on number of months of breastfeeding from a past infant and young child feeding intervention is denoted by Y_0 and follows Poisson distribution with n_0 such that:

$$Y_0 \sim Pois(\theta)$$
$$\theta \sim G(a^*, b^*)$$
$$\theta|y_0 \sim G\left(n_0\bar{y}_0 + a^*, \frac{b^*}{n_0 b^*+1}\right)$$
$$a = n_0\bar{y}_0 + a^*$$
$$b = \frac{b^*}{n_0 b^*+1}$$

If a similar intervention is carried out recently and the number of months of breastfeeding denoted by Y, is observed to follow Poisson distribution, that is:

$$Y \sim Pois(\theta)$$
$$\theta \sim G(a,b)$$
$$\theta|y \sim G\left(n\bar{y} + a, \frac{b}{nb+1}\right)$$

From the Exponential family of distribution:

$$\log(Y) = W\theta, \text{ where,}$$
$$\hat{\theta} = (W'W)^{-1}W'\log(Y)$$

and W is a matrix of covariates.

Let y_{ijt} be from Poisson distribution

$$\ell(y) = \frac{e^{y'W\theta} e^{-ne^{W\theta}}}{\prod y!} \quad (24)$$

and

$$\pi(\theta) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}\sum(\theta-\mu)^2} \quad (25)$$

$$\pi(\theta|y) = \frac{e^{y'W\theta} e^{-ne^{W\theta}}}{\prod y!} \times (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}\sum(\theta-\mu)^2} \quad (26)$$

For a particular event outcome δ_1, Suppose we have an historical data δ_0 from that same event

$$\delta_1 \in (Y_1, W_1, n_1)$$
$$\delta_0 \in (Y_0, W_0, n_0)$$

Suppose we apply Bayesian method in analyzing δ_0,

$$\pi(\theta|\delta_0) = l(\delta_0, \theta) \times \pi(\theta) \quad (27)$$

which is a power prior for θ

In analyzing the present data using the historical model in then,

$$\pi(\theta|\delta_1, \delta_0) \sim l(\theta, \delta_1) \times \pi(\theta|\delta_0) \quad (28)$$

$$\pi(\theta|\delta_1, \delta_0) \sim l(\theta, \delta_1)\left(l(\theta, \delta_0)^a \times \pi(\theta)\right) \quad (29)$$

where a is the weight of the historical data.

If one historical data is used, then $a = 1$

$$l(\theta, \delta_1) = \prod_{R_W} f(w)$$

Note

In this study, the historical data, $Y_0 \sim Pois(W_0\theta)$, and $\theta \sim N(b,B)$

$$Y_0 \sim Pois(W_0\theta)$$
$$E(Y_0|\delta_0) = e^{W_0\theta}$$
$$f(Y_0) = \frac{1}{Y_0!}\left(e^{(W_0\theta)y_0} \times e^{-e^{W_0\theta}}\right) \quad (30)$$

The likelihood is,

$$l(Y_0, \theta) = \prod_{p=0} \frac{1}{Y_0!}\left(e^{(W_0\theta)y_0} \times e^{-e^{W_0\theta}}\right) \quad (31)$$

And the prior is

$$\pi(\theta) = (2\pi\sigma^2)^{-1/2} \times e^{-\frac{1}{2\sigma^2}(\theta-b)^2} \quad (32)$$

Therefore, the posterior distribution which is the power prior is given by:

$$\pi(\theta|Y_0) = \prod_{p=0} \frac{1}{Y_0!}\left(e^{(W_0\theta)y_0} \times e^{-e^{W_0\theta}}\right)(2\pi\sigma^2)^{-1/2} \times e^{-\frac{1}{2\sigma^2}(\theta-b)^2} \quad (33)$$

where $a = 1$

For the present data δ_1 also $Y_1 \sim poisson(W\theta)$

$$f(Y_1,\theta) = \frac{1}{Y_1!}\left(e^{(W\theta)y_1} \times e^{-e^{(W\theta)}}\right) \quad (34)$$

with likelihood

$$l(Y_1,\theta) = \prod_{p=0} \frac{1}{Y_1!}\left(e^{(W\theta)y_1} \times e^{-e^{W\theta}}\right) \quad (35)$$

The posterior distribution using power prior is given by:

$$\pi(\theta|Y_1,Y_0) = \prod_{j=1} \frac{1}{Y_1!}\left(e^{(W\theta)y_1} \times e^{-e^{W\theta}}\right) \times \prod_{p=0} \frac{1}{Y_1!}\left(e^{(W_0\theta)y_0} \times e^{-e^{W_0\theta}}\right)(2\pi\sigma^2)^{-1/2} \times e^{-\frac{1}{2\sigma^2}(\theta-b)^2} \quad (36)$$

Note: $p = 1,2,\dots,n_0$, $j = 1,2,\dots,n$ and $p \neq j$

The posterior distribution is intractable since the prior is a non-conjugate prior. Estimation of the posterior distribution was carried out using Metropolis Hasting Algorithm.

2.2.2. Posterior Distribution Using Gamma Prior

The prior information $\pi(\theta)$ is from a Gamma with,

$$\theta \sim Gamma(a,b)$$

$$\pi(\theta) = \frac{1}{b^a\Gamma(a)}\theta^{b-1}e^{-\frac{\theta}{b}}$$

The posterior distribution is:

$$\pi(\theta|y,W) \propto p(y|W\theta) \times \pi(\theta)$$

That is,

$$\pi(\theta|y,W) = \prod_{j=1} \frac{1}{Y_1!}\left(e^{(W\theta)y_1} \times e^{-e^{W\theta}}\right) \times \frac{1}{b^a\Gamma(a)}\theta^{b-1}e^{-\frac{\theta}{b}}$$

Therefore, the posterior distribution for θ using $Gamma(a,b)$

$$\pi(\theta|y,W) = \frac{\theta^{b-1}}{b^a\Gamma(a)\Gamma(Y+1)}e^{-\left(\frac{\theta}{b}+e^{W\theta}-(W'\theta)n\bar{y}\right)}$$

The posterior distribution derived is intractable. Estimation of the posterior distribution will be carried out using Metropolis Hasting Algorithm

2.2.3. Posterior Distribution Using Uniform Prior (Non Informative Prior)

Recall that:

$$Y \sim Pois(\theta)$$

And

$$E(y) = V(y) = \theta$$

So, the regression model is

$$E(y) = \theta = \exp(W\theta)$$

$$p(y|\theta,W) = \frac{e^{(W\theta)y} \times e^{-e^{W\theta}}}{y!} \quad (37)$$

$$\theta \sim U(a,b)$$

$$p(\theta) = \frac{1}{b-a}$$

The posterior distribution is,

$$p(\theta|y) = \frac{e^{(W\theta)y} \times e^{-e^{W\theta}}}{y!} \times \frac{1}{b-a} \quad (38)$$

Since $U(0,1)$

$$p(\theta|y) = \frac{e^{(W\theta)y} \times e^{-e^{W\theta}}}{y!} \times 1 \quad (39)$$

which is the posterior distribution of the impact using uniform prior.

2.2.4. Posterior Distribution Using Beta Prior (Non Informative Prior)

Given that,

$$\theta \sim Beta(a,b)$$

$$\pi(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} \quad (40)$$

The posterior distribution is,

$$\pi(\theta|y) = \frac{e^{(W\theta)y} \times e^{-e^{W\theta}}}{y!} \times \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} \quad (41)$$

Since $Beta(1,1)$ and $Unif(0,1)$

$$\pi(\theta|y) = \frac{e^{(W\theta)y} \times e^{-e^{W\theta}}}{y!} \times 1 \quad (42)$$

which is the posterior distribution of the impact using beta prior.

2.3. The Bayesian Difference in Differences Estimation Approach

Define:

Y – Count variable from Poisson distribution describing the outcome of interest,

Such that:

$$Y = [Y_{10}, Y_{11}, Y_{00}, Y_{01}]$$

where:

Y_{10} – outcome of interest before intervention among those exposed to intervention (baseline)

Y_{11} – outcome of interest after intervention among those exposed to intervention (follow-up)

Y_{00} – outcome of interest before intervention among those not exposed to intervention (baseline)

Y_{01} – outcome of interest after intervention among those not exposed to intervention (follow-up

The difference in difference denoted by d is given by [1] (Lance *et al.*, 2014):

$$d = [(Y_{11} - Y_{10}) - (Y_{01} - Y_{00})]$$

where,

d – Difference in Difference

Recall that Y – count random variable from Poisson distribution describing the outcome of interest and θ is the impact parameter.

Among those exposed to intervention:

$$P(Y_{it}) = \frac{\theta_i^{y_{it}}}{y_i!} e^{-\theta_i};\ i = 0,1 \tag{43}$$

where

$$i = \begin{cases} 1 & treatment\ group \\ 0 & control\ group \end{cases}$$

At baseline,

$$P(Y_{i0}) = \frac{\theta_{i0}^{y_{i0}}}{y_{i0}!} e^{-\theta_{i0}};\ y_{i0} = 0,1,2,... \tag{44}$$

At endline,

$$P(Y_{i1}) = \frac{\theta_{i1}^{y_{i1}}}{y_{i1}!} e^{-\theta_{i1}};\ y_{i1} = 0,1,2,... \tag{45}$$

The joint distribution is given as:

$$P(Y_{i1}, Y_{i0}) = \sum_{y_{i0}=0}^{\infty} \sum_{y_{i1}=0}^{\infty} \frac{\theta_{i1}^{y_{i1}}}{y_{i1}!} e^{-\theta_{i1}} \frac{\theta_{i0}^{y_{i0}}}{y_{i0}!} e^{-\theta_{i0}} \tag{46}$$

Let $\delta_1 = Y_{11} - Y_{10}$

such that: $\delta_1 + Y_{10} = Y_{11}$

Therefore,

$$P(Y_{11} - Y_{10}) = \sum_{y_{10}=0}^{\infty} \left(\sum_{y_{11}=0}^{\delta_1 + Y_{10}} \frac{\theta_{11}^{y_{11}}}{y_{11}!} e^{-\theta_{11}} \right) \frac{\theta_{10}^{y_{10}}}{y_{10}!} e^{-\theta_{10}} \tag{47}$$

Substitute $\delta_1 + Y_{10}$ for Y_{11}

$$P(Y_{10}) = \sum_{y_{10}=0}^{\infty} \frac{\theta_1^{\delta_1 + y_{10}}}{(\delta_1 + y_{10})!} e^{-\theta_{11}} . \frac{\theta_{10}^{y_{10}}}{y_{10}!} e^{-\theta_{10}} \tag{48}$$

i.e.

$$P(Y_{10}) = e^{-\theta_{11} - \theta_{10}} \sum_{y_{10}=0}^{\infty} \theta_{11}^{\delta_1} \frac{\theta_{10}^{y_{10}} \theta_{11}^{Y_{10}}}{(\delta_1 + y_{10})! y_{10}!} \tag{49}$$

which is a modification of the Bessel function,

$$J_n(x) = \sum_{x=0}^{\infty} \frac{x^{2r+n}}{2^{2r+n}(n+r)!r!} \tag{50}$$

Hence,

$$\left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\delta_1}{2}} J_{\delta_1}\left(2\sqrt{\theta_{11}\theta_{10}}\right) = \sum_{y_{10}=0}^{\infty} (\theta_{11})^{\delta_1} \frac{(\theta_{11}\theta_{10})^{y_{10}}}{(y_{10}+d_1)! y_{10}!} dy_{10} \tag{51}$$

Therefore, the distribution of the difference in outcome before and after intervention among the intervention group is:

$$P(\delta_1) = e^{-\theta_{11} - \theta_{10}} \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\delta_1}{2}} J_{\delta_1}\left(2\sqrt{\theta_{11}\theta_{10}}\right);\ \delta_1 \tag{52}$$

which is a Skellam distribution, where,

$$J_{\delta_1}(2\sqrt{\theta_{11}\theta_{10}}) = \sum_{\delta_1=0}^{\infty} \frac{\left(\sqrt{\theta_{11}\theta_{10}}\right)^{\delta_1 + 2y_{10}}}{(\delta_1 + y_{10})!\, y_{10}!}$$

Let δ_0, denote the difference in outcome before and after intervention among the comparison group. It follows from (51) that:

$$P(\delta_0) = e^{-\theta_{01} - \theta_{00}} \left(\frac{\theta_{01}}{\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right) \tag{53}$$

where,

$$J_{\delta_0}(2\sqrt{\theta_{01}\theta_{00}}) = \sum_{\delta_1=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{\delta_0 + 2y_{00}}}{(\delta_0 + y_{00})!\, y_{00}!}$$

Therefore, if I denote difference in difference such that: $I = \delta_1 - \delta_0$ and substitute $\delta_1 = I + \delta_0$ in (52)

Then,

$$P(I) = \sum_{\delta_0=-\infty}^{\infty} \frac{\sum_{\delta_1=0}^{I+\delta_0} e^{-\theta_{11} - \theta_{10}} \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\delta_1}{2}} J_{\delta_1}\left(2\sqrt{\theta_{11}\theta_{10}}\right)}{e^{-\theta_{01} - \theta_{00}} \left(\frac{\theta_{01}}{\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)} \tag{54}$$

$$P(I) = \sum_{\delta_0=-\infty}^{\infty} \frac{e^{-\theta_{11} - \theta_{10}} \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{I+\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right)}{e^{-\theta_{01} - \theta_{00}} \left(\frac{\theta_{01}}{\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)} \tag{55}$$

$$P(I) = \sum_{\delta_0=-\infty}^{\infty} \frac{e^{-\theta_{11} - \theta_{10} - \theta_{01} - \theta_{00}} \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{I+\delta_0}{2}} \left(\frac{\theta_{01}}{\theta_{00}}\right)^{\frac{\delta_0}{2}}}{J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)} \tag{56}$$

Therefore, the distribution of the difference of differences is given by

$$P(I) = e^{-\theta_{11} - \theta_{10} - \theta_{01} - \theta_{00}} \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{I}{2}}$$

$$\sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right) \tag{57}$$

where:

$$J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) = \sum_{\delta_1=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{I+\delta_0+2y_{10}}}{(I + \delta_0 + y_{10})!\, y_{10}!}$$

and

$$J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right) = \sum_{\delta_0=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{\delta_0+2y_{00}}}{(\delta_0 + y_{00})!\, y_{00}!}$$

Hence, the distribution of the difference in difference for count outcome from Poisson distribution is a 4-Parameter Skellam distribution with mean

$$E(I) = \theta_{11} - \theta_{10} - \theta_{01} + \theta_{00}$$

And variance

$$V(I) = \theta_{11} + \theta_{10} + \theta_{01} + \theta_{00}$$

The likelihood of I

$$L(I) = \prod_{j=1}^{n} I_{ijt} \tag{58}$$

$$L(I) = \prod_{j=1}^{n} \left[\left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}} e^{-\theta_{11} - \theta_{10} - \theta_{01} - \theta_{00}} \times \sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right) \right] \tag{59}$$

$$L(I) = \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}} e^{-\theta_{11}-\theta_{10}-\theta_{01}-\theta_{00}}$$

$$\ln \sum_{\delta_0=-\infty}^{\infty} \left[\left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)\right] \quad (60)$$

For

$$J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) = \sum_{\delta_1=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{I+\delta_0+2y_{10}}}{(I+\delta_0+y_{10})!\, y_{10}!}$$

and

$$J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right) = \sum_{\delta_0=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{\delta_0+2y_{00}}}{(\delta_0+y_{00})!\, y_{00}!}$$

Therefore,

$$L(I) = \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}} e^{-\theta_{11}-\theta_{10}-\theta_{01}-\theta_{00}} \sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}}$$

$$L(I) = \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}} e^{-\theta_{11}-\theta_{10}-\theta_{01}-\theta_{00}} \sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}}$$

$$\sum_{\delta_1=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{I+\delta_0+2y_{10}}}{(I+\delta_0+y_{10})!y_{10}!} \times \sum_{\delta_0=0}^{\infty} \frac{\left(\sqrt{\theta_{01}\theta_{00}}\right)^{\delta_0+2y_{00}}}{(\delta_0+y_{00})!y_{00}!}$$

2.3.1. Posterior distribution using Gamma Prior

The prior distribution for the $\theta \sim Gamma(a_{it}, b_{it})$

$$\pi(\theta_{it}) = \frac{\theta_{it}}{b_{it}^{a_{it}}\Gamma(a_{it})} e^{-\frac{\theta_{it}}{b_{it}}} \quad (61)$$

and, these are:

$$\pi(\theta_{11}) = \frac{\theta_{11}}{b_{11}^{a_{11}}\Gamma(a_{11})} e^{-\frac{\theta_{11}}{b_{11}}} \quad (62)$$

$$\pi(\theta_{10}) = \frac{\theta_{10}}{b_{10}^{a_{10}}\Gamma(a_{10})} e^{-\frac{\theta_{10}}{b_{10}}} \quad (63)$$

$$\pi(\theta_{01}) = \frac{\theta_{01}}{b_{01}^{a_{01}}\Gamma(a_{01})} e^{-\frac{\theta_{01}}{b_{01}}} \quad (64)$$

$$\pi(\theta_{00}) = \frac{\theta_{00}}{b_{00}^{a_{00}}\Gamma(a_{00})} e^{-\frac{\theta_{00}}{b_{00}}} \quad (65)$$

The joint prior distribution is:

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) = \pi(\theta_{11})\pi(\theta_{10})\pi(\theta_{01})\pi(\theta_{00}) \quad (66)$$

Given that,

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\, \pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) \quad (67)$$

Such that

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) = \frac{\theta_{11}\theta_{10}\theta_{01}\theta_{00}}{b_{11}^{a_{11}}b_{10}^{a_{10}}b_{01}^{a_{01}}b_{00}^{a_{00}}} \times \frac{e^{-\frac{\theta_{11}}{b_{11}}}e^{-\frac{\theta_{10}}{b_{10}}}e^{-\frac{\theta_{01}}{b_{01}}}e^{-\frac{\theta_{00}}{b_{00}}}}{\Gamma(a_{11})\Gamma(a_{10})\Gamma(a_{01})\Gamma(a_{00})} \quad (68)$$

and the posterior distribution is:

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) = n\left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}} e^{-n(\theta_{11}+\theta_{10}+\theta_{01}+\theta_{00})}$$

$$\times \sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)$$

$$\times\, \theta_{11}\theta_{10}\theta_{01}\theta_{00} e^{-\left(\frac{\theta_{11}}{b_{11}}+\frac{\theta_{10}}{b_{10}}+\frac{\theta_{01}}{b_{01}}+\frac{\theta_{00}}{b_{00}}\right)}$$

$$\times\, \theta_{11}\theta_{10}\theta_{01}\theta_{00} e^{-\left(\frac{\theta_{11}}{b_{11}}+\frac{\theta_{10}}{b_{10}}+\frac{\theta_{01}}{b_{01}}+\frac{\theta_{00}}{b_{00}}\right)} \quad (69)$$

The posterior distribution derived is intractable. Estimation of the posterior distribution will be carried out using Metropolis Hasting Algorithm.

2.3.2. Posterior Distribution Using Normal Prior

Given that:

$$\theta_{it} \sim N(\tilde{\theta}_{it}, \tau_{it}) \quad (70)$$

The prior distribution:

$$\pi(\theta_{it}) = (2\pi\tau_{it}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{it}^2}(\theta_{it}-\tilde{\theta}_{it})^2} \quad (71)$$

is given by:

$$\pi(\theta_{11}) = (2\pi\tau_{11}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{11}^2}(\theta_{11}-\tilde{\theta}_{11})^2} \quad (72)$$

$$\pi(\theta_{10}) = (2\pi\tau_{10}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{10}^2}(\theta_{10}-\tilde{\theta}_{10})^2} \quad (73)$$

$$\pi(\theta_{01}) = (2\pi\tau_{01}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{01}^2}(\theta_{01}-\tilde{\theta}_{01})^2} \quad (74)$$

$$\pi(\theta_{00}) = (2\pi\tau_{00}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{00}^2}(\theta_{00}-\tilde{\theta}_{00})^2} \quad (75)$$

The joint prior distribution is

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) = \pi(\theta_{11})\pi(\theta_{10})\pi(\theta_{01})\pi(\theta_{00}) \quad (76)$$

Given that,

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) = (2^4\pi^4\tau_{11}^2\tau_{10}^2\tau_{01}^2\tau_{00}^2)^{-\frac{1}{2}} e^{-\frac{1}{2}\left[\left(\frac{\theta_{11}-\tilde{\theta}_{11}}{\tau_{11}}\right)^2+\left(\frac{\theta_{10}-\tilde{\theta}_{10}}{\tau_{10}}\right)^2+\left(\frac{\theta_{01}-\tilde{\theta}_{01}}{\tau_{01}}\right)^2+\left(\frac{\theta_{00}-\tilde{\theta}_{00}}{\tau_{00}}\right)^2\right]} \quad (77)$$

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\, \pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) \quad (78)$$

Such that:

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto n\left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}}$$

$$\times\, e^{-n(\theta_{11}+\theta_{10}+\theta_{01}+\theta_{00})+\frac{1}{2}\left[\left(\frac{\theta_{11}-\tilde{\theta}_{11}}{\tau_{11}}\right)^2+\left(\frac{\theta_{10}-\tilde{\theta}_{10}}{\tau_{10}}\right)^2+\left(\frac{\theta_{01}-\tilde{\theta}_{01}}{\tau_{01}}\right)^2+\left(\frac{\theta_{00}-\tilde{\theta}_{00}}{\tau_{00}}\right)^2\right]}$$

$$\times \sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right) \quad (79)$$

The posterior distribution derived is intractable. Estimation of the posterior distribution was carried out using Metropolis Hasting Algorithm.

2.3.3. Posterior Distribution Using Power Prior

In this case, the historical data information I_0 with likelihood, $L(I_0|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$. Therefore, the prior distribution for the multiple of likelihood of the historical data $[I_0|\theta_{it}]^a$ where a is the weights of the historical data which may be determined by different number of data historical data available and $\pi(\theta_{it})$ prior information of the parameter [12]. If the historical data is from the same distribution as current data and the prior information of the parameter follows:

$$\theta_{it} \sim N(\tilde{\theta}_{it}, \tau_{it})$$

$$L(\theta_{it}|I_0) \propto \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I_0}{2}} \times$$

$$\left[\sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)\right]$$

$$\pi(\theta_{it}) = (2\pi\tau_{it}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{it}^2}(\theta_{it}-\tilde{\theta}_{it})^2} \quad (80)$$

and, these are:

$$\pi(\theta_{11}) = (2\pi\tau_{11}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{11}^2}(\theta_{11}-\tilde{\theta}_{11})^2} \quad (81)$$

$$\pi(\theta_{10}) = (2\pi\tau_{10}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{10}^2}(\theta_{10}-\tilde{\theta}_{10})^2} \quad (82)$$

$$\pi(\theta_{01}) = (2\pi\tau_{01}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{01}^2}(\theta_{01}-\tilde{\theta}_{01})^2} \quad (83)$$

$$\pi(\theta_{00}) = (2\pi\tau_{00}^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau_{00}^2}(\theta_{00}-\tilde{\theta}_{00})^2} \quad (84)$$

The joint prior distribution is

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto$$
$$L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\pi(\theta_{11})\pi(\theta_{10})\pi(\theta_{01})\pi(\theta_{00}) \quad (85)$$

For $a = 1$, that is using only one historical data, we have

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto$$
$$L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\, \pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}) \quad (86)$$

and therefore, the power prior is as follows:

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I_0}{2}}$$
$$\times \left[\sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)\right]$$
$$\times (2^4\pi^4\tau_{11}^2\tau_{10}^2\tau_{01}^2\tau_{00}^2)^{-\frac{1}{2}}$$
$$e^{-\frac{1}{2}\left[\left(\frac{\theta_{11}-\tilde{\theta}_{11}}{\tau_{11}}\right)^2+\left(\frac{\theta_{10}-\tilde{\theta}_{10}}{\tau_{10}}\right)^2+\left(\frac{\theta_{01}-\tilde{\theta}_{01}}{\tau_{01}}\right)^2+\left(\frac{\theta_{00}-\tilde{\theta}_{00}}{\tau_{00}}\right)^2\right]} \quad (87)$$

and the posterior distribution using the power prior (230) is as follows:

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto$$
$$L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\, \pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$$

Where

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto$$
$$L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\, \pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$$

So,

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$$
$$\times \left(L(I|\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\, \pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})\right)$$

The posterior for the impact for count using the power prior is:

$$\pi(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}|I) \propto \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I}{2}}$$
$$\times \left[\sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)\right]$$
$$\times \left(\frac{\theta_{11}}{\theta_{10}}\right)^{\frac{\sum I_0}{2}}$$
$$\times \left[\sum_{\delta_0=-\infty}^{\infty} \left(\frac{\theta_{11}\theta_{01}}{\theta_{10}\theta_{00}}\right)^{\frac{\delta_0}{2}} J_{I+\delta_0}\left(2\sqrt{\theta_{11}\theta_{10}}\right) J_{\delta_0}\left(2\sqrt{\theta_{01}\theta_{00}}\right)\right] \quad (88)$$

The posterior distribution derived is intractable. Estimation of the posterior distribution was carried out using Metropolis Hasting Algorithm.

2.4. Data and Prior Information

Simulated and real-life data were used in this study. Prior information was determined based on literature search.

Using simulated data, the outcome variable (Y) was simulated as a count outcome data type from Poisson distribution, with sample size of 40. The Bayesian estimation was carried out using Metropolis-Hastings algorithm. The posterior distributions considered are based on Gamma prior. Using 10000 iterations, the posterior mean, bias and Mean Squared Error (MSE) are reported. The report was recorded at various prior parameters (a, b), where;

$a = 0.001, 0.01, 0.1\ and\ 1$ and b= $0.01, 0.1\ and\ 1$. The parameters used for the simulations of count data are from Poisson distribution with parameter $\theta = 10$ for program group at follow-up, Poisson distribution with parameter $\theta = 4.5$ for program group at baseline, Poisson distribution with parameter $\theta = 5.5$ for comparison group at follow-up and Poisson distribution with parameter $\theta = 4$ for comparison group baseline.

Using real life data, the posterior distribution for Poisson distributed count response data was estimated using data from an Infant and Young Child Feeding and Nutrition intervention in North East Nigeria. Data was available number of months of breastfeeding from initiation to termination. The response variable is the number of months of breastfeeding among those who participated in a Breastfeeding Support Group Program (BFSG) and those who did not. Mothers were followed up for at least two years from delivery of the reference child. Breastfeeding was defined as Breast milk (including milk expressed or from a wet nurse given to a child on demand [13]. Controls were selected among mothers in different communities who did not participate in the breastfeeding support program. The objective of the study was to determine the impact of breastfeeding support program on duration of breastfeeding among mothers who participated in breastfeeding support groups and those who did not.

2.5. Prior Information on Breastfeeding Interventions

The WHO recommended breastfeeding from age 0 to 23 months as part of the Infant and Young Child Feeding Policy. However, several challenges prevent breastfeeding in Nigeria. Hence, several interventions are being carried out to increase the duration of breastfeeding. Many of these programs often were only able to improve overall duration of breastfeeding by a small percentage. In a systematic review to determine effective approaches to Social and Behavior Change programs for reducing stunting and Anemia [14].

From this review, the following studies were analyzed and formed basis for determining the hyperparameter for the prior distribution [14]. A study was conducted among breastfeeding mothers in Vietnam on impact of Counseling in homes and group education in communities on number of months of breastfeeding [15]. The study revealed an increase in the mean duration of breastfeeding from 16.6 before the intervention to 16.8 after intervention. Also, another study provided result of impact evaluation of a breastfeeding counselling and support group program in Turkey [16]. The program increased duration of breastfeeding from 17.83 months before intervention to 21.17 after intervention. As presented in another study conducted to determine impact of breastfeeding counselling in homes and health facilities in Turkey, the mean duration of breastfeeding increased from 12.1 months before the intervention to 15.1 months after the intervention [17]. Based on the result, a range of values for the hyper-parameter was determined as ($0<\theta<1$), with specific point estimate of the mean at 0.24, representing 24% increase in number of months of breastfeeding [17].

2.6. Statistical Data Analysis

Estimation of Posterior distribution was carried out using R, a statistical computing and graphics software developed by Bell's laboratory based on the 'S' system. This software is widely used among statisticians and is considered the most robust statistical software for statistical analysis. A standard result of a closed form posterior distribution exists for the Poisson model without covariates. However, no conjugate prior exists for the ($k \times 1$) parameter vector β in the Poisson regression model. Hence parameter estimation will be conducted by evaluation of the exact posterior distribution using the Metropolis Hasting Algorithm [18]. The Metropolis Hasting Algorithm is a special type of Markov chain Monte Carlo (MCMC) approach that allows for the estimation of posterior distributions in Bayesian statistics through simulation [18].

2.7. Comparison of Result from Bayesian and Classical Procedures

In Statistics, Mean Squared Error (MSE) is defined as Mean or Average of the square of the difference between actual and estimated values. The mean square error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is the function of θ defined by $E\left(\hat{\theta}-\theta\right)^2$. The MSE measures the average squared difference between the estimator $\hat{\theta}$ and the parameter $\hat{\theta}$, which is considered a reasonable measure of performance for an estimator. MSE is used to check how close estimates are to actual values. Hence, the lower the MSE, the closer is the estimate to the actual. The most common risk function used for Bayesian estimation is the mean square error. MSE is a useful tool for comparing estimate from Bayesian and classical procedures [19-21].

3. Result

3.1. Result from Simulated data

Using simulated data, the mean of the posterior distributions at different values of $\tilde{\theta}$ were computed. This section presents the result of analysis. Decision is based on Minimum Mean Squared Error

Table 1. Result of classical approach (Poisson regression)

$\hat{\theta}$	MSE	Lower CI	Upper CI
0.1718	0.4680	0.1429	0.2006

Table 2. Bayesian regression with Gamma Prior

$\tilde{\theta}$	$\hat{\theta}$	Impact	MSE	Lowe CI	Upper CI
0.1	0.0824	0.0859	0.0065	0.0001	0.2559
0.2	0.0754	0.0783	0.0068	0.0004	0.2376
0.24	0.0169	0.0170	0.0014	0.0002	0.0546
0.3	0.0885	0.0925	0.0068	0.0003	0.2564
0.4	0.1034	0.1089	0.0057	0.0020	0.2427
0.5	0.1077	0.1137	0.0061	0.0014	0.2580
0.6	0.0947	0.0993	0.0045	0.0017	0.2198
0.7	0.0976	0.1025	0.0060	0.0013	0.2387
0.8	0.0997	0.1048	0.0077	0.0032	0.2460
0.9	0.1007	0.1059	0.0090	0.0027	0.2187
1.0	0.1067	0.1126	0.0140	0.0077	0.2369

Table 3. Result of Bayesian regression with Power prior

$\tilde{\theta}$	$\hat{\theta}$	Impact	MSE	Lower CI	Upper CI
0	0.0376	0.0383	0.0034	0.0015	0.1227
0.1	0.0371	0.0378	0.0057	0.0008	0.1128
0.2	0.0342	0.0348	0.0036	0.0012	0.1166
0.24	0.0316	0.0321	0.0031	0.0014	0.0990
0.3	0.0354	0.0360	0.0034	0.0019	0.1060
0.4	0.0359	0.0366	0.0029	0.0004	0.1149
0.5	0.0375	0.0382	0.0017	0.0010	0.1081
0.6	0.0364	0.0371	0.0023	0.0006	0.1100
0.7	0.0336	0.0342	0.0023	0.0019	0.0977
0.8	0.0328	0.0333	0.0037	0.0012	0.0934
0.9	0.0383	0.0390	0.0071	0.0010	0.0953
1.0	0.0420	0.0429	0.0087	0.0036	0.1026

Table 4. Result of Bayesian estimation with Gamma prior

$\tilde{\theta}$	$\hat{\theta}$	MSE	Lower CI	Upper CI
0.1	0.1139	0.0086	0.0043	0.3473
0.2	0.1147	0.0087	0.0042	0.3498
0.24	0.1142	0.0087	0.0045	0.3482
0.3	0.1141	0.0087	0.0045	0.3481
0.4	0.1124	0.0085	0.0040	0.3449
0.5	0.1140	0.0086	0.0042	0.3460
0.6	0.1131	0.0085	0.0042	0.3463
0.7	0.1148	0.0088	0.0043	0.3509
0.8	0.1127	0.0084	0.0042	0.3471
0.9	0.1138	0.0087	0.0043	0.3488
1.0	0.1134	0.0085	0.0042	0.3453

Table 5. Result of Bayesian estimation with Power prior

$\tilde{\theta}$	$\hat{\theta}$	MSE	Lowe CI	Upper CI
0.1	0.2282	0.0343	0.0088	0.6931
0.2	0.2294	0.0346	0.0091	0.7003
0.24	0.2274	0.0339	0.0088	0.6876
0.3	0.2296	0.0349	0.0087	0.7005
0.4	0.2269	0.0340	0.0087	0.6903
0.5	0.2299	0.0354	0.0090	0.7051
0.6	0.2307	0.0356	0.0092	0.7062
0.7	0.2281	0.0345	0.0083	0.6958
0.8	0.2283	0.0343	0.0086	0.6945
0.9	0.2296	0.0348	0.0089	0.7027
1.0	0.2279	0.0346	0.0082	0.6957

Given that the true value for the impact parameter (θ) from simulated data is 0.25, it was observed that the result derived from the proposed Bayesian DID estimation is closest to the true value compared to the existing classical approach as presented in table 6 and figure 1.

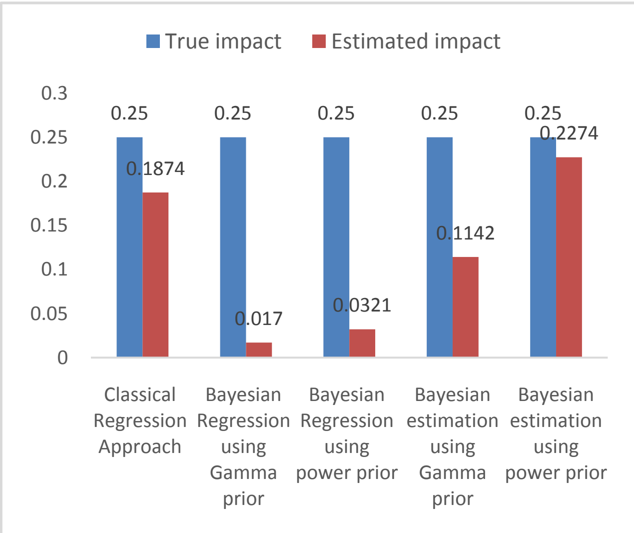

Figure 1. Comparison of true value and estimated value of impact based on proposed and existing approaches for count outcome using simulated data

Table 6. Comparison between result from classical and Bayesian procedures for count outcome based on simulated data

Method	Impact	MSE
Classical Regression Approach	0.1874	0.4680
Bayesian Regression Using Gamma prior	0.0170	0.0014
Bayesian Regression Using Power Prior	0.0321	0.0031
Bayesian Estimation Using Gamma Prior	0.1142	0.0087
Bayesian Estimation Using Power Prior	0.2274	0.0339

3.2. Result from Real Life Data

Table 7. Summary of data

Statistic	n	Mean	Variance
Control site	72	13	12
Treatment site	68	24.4	24.5

From table 7, the data suggest the outcome variable may have come from a Poisson distribution evidenced by the similarity between the mean and variance. Hence a Poisson regression analysis was carried out. Table 8 provide result from the logistic regression.

Table 8. Result from the classical approach (Poisson regression)

θ	Impact	MSE	P value	Lower CL	Upper CL
0.4147	0.5139	0.0616	<0.001	0.2975	0.5324

The deviance statistic and the Pearson statistic are provided in table 9.

Table 9. Result of goodness of fit test

Deviance	P value	Pearson	P value
222.1541	0.9999	210.5973	1.000

Result of the goodness of fit test revealed that the model fit the data suggesting that data used is from Poisson distribution. The impact estimate is obtained as 0.4147, with 0.0616 MSE, and observed to be significant ($p<0.001$) at 5% level of significance. This implies that participation in Breastfeeding Support Group (BFSG) increases number of months of breastfeeding by about 51.3% ($P<0.001$).

Table 10. Result of Bayesian regression using Gamma prior

$\tilde{\theta}$	$\hat{\theta}$	Impact	MSE	Lower CI	Upper CI
0.1	0.4234	0.5271	0.0058	0.2821	0.5660
0.2	0.4214	0.5241	0.0054	0.2829	0.5561
0.24	0.4244	0.5287	0.0058	0.2850	0.5659
0.3	0.4230	0.5265	0.0055	0.2854	0.5534
0.3	0.4219	0.5249	0.0062	0.2852	0.5535
0.4	0.4191	0.5206	0.0066	0.2792	0.5626
0.5	0.4184	0.5195	0.0060	0.2783	0.5528
0.6	0.4216	0.5244	0.0061	0.2911	0.5513
0.7	0.4189	0.5203	0.0070	0.2862	0.5492
0.8	0.4185	0.5197	0.0050	0.2857	0.5467
0.9	0.4174	0.5180	0.0050	0.2869	0.5340
1.0	0.4159	0.5157	0.0055	0.2933	0.5446

3.3. Result of Bayesian Regression Using Gamma Prior

Bayesian analysis was carried out using Gamma prior. Table 10 provides result of the analysis.

The impact estimate obtained at $\hat{\theta}= 0.24$ is 0.4244, with MSE 0.0058. This implies that participation in Breastfeeding Support Group (BFSG) increases number of months of breastfeeding by about 53% over baseline values.

3.4. Result of Bayesian Regression Using Power Prior

Bayesian analysis was carried out using power prior. Historical data for the power prior distribution was obtained from a similar study conducted in Borno state in 2018 among a different population. Table 11 provides result of the analysis.

Table 11. Result of Bayesian analysis using power prior

$\tilde{\theta}$	$\hat{\theta}$	Impact	MSE	Lower CI	Upper CI
0	0.3389	0.4034	0.0087	0.2458	0.4532
0.1	0.3355	0.3986	0.0100	0.2162	0.4543
0.2	0.3423	0.4082	0.0077	0.2327	0.4609
0.24	0.3385	0.4028	0.0072	0.2453	0.4664
0.3	0.3375	0.4014	0.0090	0.2440	0.4549
0.4	0.3399	0.4048	0.0086	0.2182	0.4533
0.5	0.3411	0.4065	0.0062	0.2417	0.4541
0.6	0.3414	0.4069	0.0063	0.2436	0.4537
0.7	0.3426	0.4086	0.0051	0.2330	0.4561
0.8	0.3411	0.4065	0.0043	0.2315	0.4499
0.9	0.3441	0.4107	0.0048	0.2487	0.4519
1.0	0.3432	0.4095	0.0055	0.2313	0.4526

The impact estimate obtained at θ = 0.24 is 0.3385, with MSE 0.0072. This implies that participation in Breastfeeding Support Group (BFSG) increases number of months of breastfeeding by about 40% over baseline values.

3.5. Result of Bayesian Estimation Using Gamma Prior

Bayesian estimation was carried out using Gamma prior. Table 12 provides result of the analysis.

Table 12. Result of Bayesian estimation using Gamma prior

$\tilde{\theta}$	$\hat{\theta}$	Impact	MSE	Lower CI
0.1	0.4353	0.0061	0.2787	0.5913
0.2	0.4356	0.0061	0.2788	0.5929
0.24	0.4354	0.0061	0.2781	0.5925
0.3	0.4356	0.0062	0.2768	0.5939
0.4	0.4357	0.0062	0.2776	0.5935
0.5	0.4354	0.0062	0.2764	0.5952
0.6	0.4354	0.0061	0.2765	0.5926
0.7	0.4354	0.0062	0.2789	0.5927
0.8	0.4354	0.0061	0.2790	0.5926
0.9	0.4356	0.0061	0.2791	0.5928
1.0	0.4356	0.0062	0.2770	0.5929

The impact estimate is obtained at $\hat{\theta}$ = 0.24 is 0.4354, with MSE 0.0061. This implies that participation in Breastfeeding Support Group (BFSG) increases number of months of breastfeeding by about 44% over baseline values.

3.6. Result of Bayesian Estimation Using Power Prior

Bayesian analysis was carried out using power prior. Historical data for the power prior distribution was obtained from a similar study conducted in Borno state in 2018 among a different population. Table 13 provides result of the analysis.

The impact estimate is obtained at $\hat{\theta}$ = 0.24 is 0.4476, with MSE 0.0061. This implies that participation in Breastfeeding Support Group (BFSG) increases number of months of breastfeeding by about 45% over baseline values.

Table 13. Result of Bayesian estimation using power prior

$\tilde{\theta}$	$\hat{\theta}$	Impact	MSE	Lower CI
0.1	0.4474	0.0060	0.2909	0.6018
0.2	0.4476	0.0059	0.2930	0.6032
0.24	0.4476	0.0061	0.2907	0.6055
0.3	0.4475	0.0060	0.2919	0.6034
0.4	0.4475	0.0060	0.2916	0.6047
0.5	0.4476	0.0060	0.2929	0.6015
0.6	0.4478	0.0060	0.2929	0.6048
0.7	0.4476	0.0060	0.2921	0.6035
0.8	0.4479	0.0060	0.2918	0.6031
0.9	0.4473	0.0060	0.2910	0.6010
1.0	0.4476	0.0061	0.2800	0.6050

3.7. Comparison of Result from Classical and Bayesian Procedure

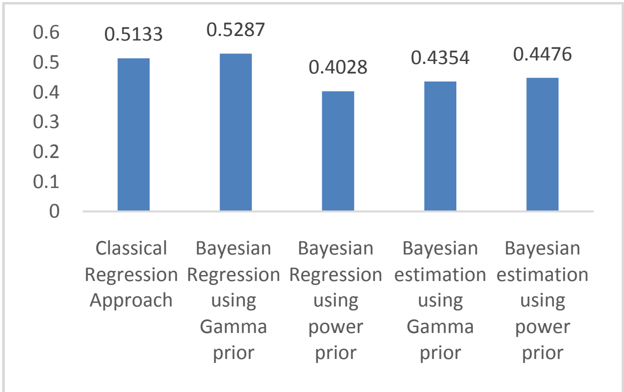

Figure 2. Comparison of impact estimates for count outcome based on proposed and existing approaches using real life data

Table 14 shows comparison of result obtained from the proposed Bayesian and existing classical approaches. The "Bayesian DID estimation approach provided a realistic impact estimate and lower Mean Squared Error compared to the existing classical approach. Based on this approach, it can be inferred that exposure to BFSG increases number of months of breastfeeding by 45% compared to 51% realized from the existing classical method, as presented in table 14 and figure 2.

Table 14. Comparison between result from classical and Bayesian procedures

Method	Impact	MSE
Classical Regression Approach	0.5133	0.0616
Bayesian Regression using Gamma prior	0.5287	0.0058
Bayesian Regression using power prior	0.4028	0.0072
Bayesian Estimation Using Gamma prior	0.4354	0.0061
Bayesian Estimation Using Power Prior	0.4476	0.0061

3.8. Result from Non-Informative Prior (Uniform and Beta priors)

Using non informative prior, the results obtained from

Bayesian regression and Bayesian estimation approaches were very close to the result from the classical regression approach. However, the proposed Bayesian estimation of the distribution of the DID approach produced a lower Mean Squared Error, compared to the existing classical method, as presented in table 15.

Table 15. Comparison of Bayesian and Classical approach based on real life data using non informative prior

Method	$\hat{\theta}$	MSE
Classical Regression Approach	0.1874	0.4680
Bayesian Estimation approach with Beta Prior	0.1199	0.0085
Bayesian Estimation Approach with Uniform Prior	0.1199	0.0085

4. Conclusions

Estimating Impact of experimental intervention from the classical approach is widely adopted among evaluators and researchers. Bayesian approach are being newly introduced into experimental evaluation and this study provides framework for adopting Bayesian approach in experimental intervention when outcome of interest is count and from Poisson distribution. In adopting the Bayesian approach, findings from this study showed that Bayesian parameter estimation produced lower impact estimate and lower Mean Squared Error compared to the Bayesian regression and classical regression approaches. Adopting a procedure that provides a more precise estimate of impact of experimental intervention will help normalize outrageous claims of intervention efficacy. Findings from this study will provide opportunity to apply a better impact estimation procedure than what is commonly adopted to improve decision from experimental intervention.

ACKNOWLEDGEMENTS

The authors would like to acknowledge faculty staff of the Department of Statistics, University of Ilorin, Ilorin, Nigeria for the review of the study.

REFERENCES

[1] Lance P, Guilkey D, Hattori A, Angeles G. How do we know if a program made a difference? A guide to statistical methods for program impact evaluation [Internet]. Chapel Hill, North Carolina: MEASURE Evaluation; 2014. Available from: https://www.measureevaluation.org/resources/publications/ms-14-87-en.html.

[2] Least squares, maximum likelihood, and Bayesian parameter estimation. In: Genomic Selection in Animals. Hoboken, NJ: John Wiley & Sons, Inc; 2016. p. 21–9.

[3] Adebayo SB. Bayesian geoadditive modelling of breastfeeding initiation in Nigeria. J Appl Econ (Chichester Engl) [Internet]. 2004; 19(2): 267–81. Available from: http://dx.doi.org/10.1002/jae.732.

[4] Gill J. Bayesian methods: A social and behavioral sciences approach, third edition. 3rd ed. Philadelphia, PA: Chapman & Hall/CRC; 2014.

[5] Jackman S. Bayesian analysis for the social sciences. Hoboken, NJ: Wiley-Blackwell; 2009.

[6] Ipadeola O, Jolayemi E. Bayesian estimation of impact in experimental interventions with continuous outcome. Int J Stat Appl Math [Internet]. 2021; 6(1): 223–9. Available from: http://dx.doi.org/10.22271/maths.2021.v6.i1c.652.

[7] Feng C, Wang H, Lu N, Chen T, He H, Lu Y, et al. Log-transformation and its implications for data analysis. Shanghai Arch Psychiatry [Internet]. 2014; 26(2): 105–9. Available from: http://dx.doi.org/10.3969/j.issn.1002-0829.2014.02.009.

[8] Comoglio F, Fracchia L, Rinaldi M. Bayesian inference from count data using discrete uniform priors. PLoS One [Internet]. 2013; 8(10): e74388. Available from: http://dx.doi.org/10.1371/journal.pone.0074388.

[9] Morita S, Thall PF, Müller P. Evaluating the impact of prior assumptions in Bayesian biostatistics. Stat Biosci [Internet]. 2010; 2(1): 1–17. Available from: http://dx.doi.org/10.1007/s12561-010-9018-x.

[10] Neelon BH, O'Malley AJ, Normand S-LT. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. Stat Modelling [Internet]. 2010; 10(4): 421–39. Available from: http://dx.doi.org/10.1177/1471082X0901000404.

[11] Kaciroti NA, Raghunathan TE, Schork MA, Clark NM. A Bayesian model for longitudinal count data with non-ignorable dropout. J R Stat Soc Ser C Appl Stat [Internet]. 2008; 57(5): 521–34. Available from: http://dx.doi.org/10.1111/j.1467-9876.2008.00628.x.

[12] Chen M-H, Ibrahim JG. Power prior distributions for Regression Models. Statistical Science. 2000; 15(1). Available from: 10.1214/ss/1009212673.

[13] Noel-Weiss J, Boersma S, Kujawa-Myles S. Questioning current definitions for breastfeeding research. International Breastfeeding Journal. 2012; 7(1). Available from: 10.1186/1746-4358-7-9.

[14] Lamstein S, Anson M, Beall K, Williams T, Collaiezzi B, Aakesson A, et al. Evidence of effective approaches to social and behavior change communication for preventing and reducing stunting and anemia [Internet]. SPRING. 2014. Available from: https://www.spring-nutrition.org/publications/series/evidence-effective-approaches-social-and-behavior-change-communication.

[15] Mackintosh UA, Marsh DR, Schroeder DG. Sustained positive deviant child care practices and their effects on child growth in Viet Nam. Food and Nutrition Bulletin. 2002; 23(4_suppl_1): 16–25.

[16] Duyan Çamurdan A, Özkan S, Yüksel D, Pasli F, Şahin F, Beyazova U. The effect of the baby-friendly hospital initiative on long-term breast feeding. International Journal of Clinical Practice. 2007; 61(8): 1251–5.

[17] Aksu H, Küçük M, Düzgün G. The effect of postnatal

breastfeeding education/support offered at home 3 days after delivery on breastfeeding duration and knowledge: A randomized trial. The Journal of Maternal-Fetal & Neonatal Medicine. 2010; 24(2): 354–61.

[18] Chib S, Greenberg E. Understanding the metropolis-hastings algorithm. The American Statistician. 1995; 49(4): 327–35.

[19] Singh S, Arabi Belaghi R, Noori Asl M. Estimation and prediction using classical and Bayesian approaches for Burr III model under Progressive Type-I hybrid censoring. International Journal of System Assurance Engineering and Management. 2019; 10(4): 746–64.

[20] Ennouri K, Ben Ayed R, Mazzarello M, Ottaviani E, Hertelli F, Azzouz H. Classical and bayesian predictions applied to bacillus toxin production. 3 Biotech. 2016; 6(2).

[21] Oliphant, Travis E. A Bayesian perspective on estimating mean, variance, and standard deviation from data. Faculty Publications. 2006; 278. https://scholarsarchive.byu.edu/facpub/278.