Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:127
https://doi.org/10.28919/cmbn/9440
ISSN: 2052-2541

COMPARISON OF BAYESIAN DYNAMIC LINEAR AND MIXED EFFECTS INTERRUPTED TIME SERIES MODELS IN ASSESSING UGANDA'S NEONATAL MORTALITY REDUCTION SINCE THE SDGS ONSET

BAMWEBAZE GEORGE[1,*], GICHUHI A. WAITITU[2], RICHARD O. AWICHI[3]

[1]Department of Mathematics and Data Science, Pan African University Institute for Basic Science, Technology and Innovation, Nairobi, Kenya

[2]Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

[3]Department of Science and Vocational Education, Lira University, Lira, Uganda

Abstract. The study makes a comparison of two models used to assess Uganda's effort in reducing neonatal mortality since the onset of the Sustainable Development Goals. These two models are Bayesian Dynamic Linear and Mixed Effects Interrupted Time Series Models. The study made use of secondary data obtained from the country's Ministry of Health spanning from January 2015 to December 2023. To determine the most appropriate model, the study conducted both Robust tests of model diagnosis and K-fold cross-validation tests as a global measure of model comparison. The Bayesian Dynamic Linear Model outperformed the Mixed Effects Interrupted Time Series Model on three out of the four criteria considered to assess the two models. Besides having a higher Cross-Validated Root Mean Square Error, the Bayesian Dynamic Linear Model does better on robust tests and has a higher Cross-Validated Log Likelihood. Therefore, the Bayesian Dynamic Linear Model outshone Mixed Effects Interrupted Time Series Model as a preferred model for Neonatal mortality analysis amidst SDGs onset impact evolution, in alignment with the study's methodological innovations and overarching research objectives.

*Corresponding author
E-mail address: georgebamwebaze@gmail.com
Received July 15, 2025

Keywords: neonatal mortality; Bayesian dynamic linear model; mixed effects interrupted time series model; K-fold cross validation.

2020 AMS Subject Classification: 65C20, 92B15.

1. INTRODUCTION

Uganda has made substantial investments to improve maternal and child health [9, 12]. Despite this, neonatal mortality (NM) remains a significant concern in the country [11]. Like any other country, Uganda prefers to have a healthy economy to achieve a wealthy one. The 2016 launch of the Sustainable Development Goals (SDGs) emphasized a global reduction of neonatal mortality (NNM) to fewer than 12 deaths per 1000 live births by 2030 [10]. This study evaluates the impact of the SDGs on NNM in Uganda using an Interrupted Time Series (ITS) and Bayesian approaches. These are well-established methods for causal inference in observational health data [2, 14].

This study extends the standard Interrupted Time Series (ITS) design by incorporating mixed effects modeling whereas traditional ITS assumes homogeneity and is often applied to aggregate data [14, 2]. At the same time, the mixed effects ITS (ME-ITS) model accounts for the hierarchical structure of the data, the variability between groups in baseline levels and trends. This allows a more accurate estimation of the effects of the intervention in the presence of clustering and heterogeneity [8].

2. METHODS

2.1. Data Types, Sources, and Timeframe. The study used data from the DHIS2 (District Health Information System 2) obtained from the Ministry of Health, Uganda.

The main outcome assessed by the study was the monthly neonatal mortality at the district levels in Uganda from January 2015 to December 2023; hence, the study contained 15,168 observations.

2.2. Study Variables.

TABLE 1. List of study variables

Variable	Variable name	Description	Variable nature
v01	Neonatal	The monthly number of newborn children dying within their first 28 days of life per district.	Discrete
v02	Year	The calendar year for which the data was recorded.	Ordinal
v03	Month	The month of the calendar year for which the data was recorded.	Ordinal
v04	District	Name of the district from which the data was recorded.	Nominal
v05	Region	The administrative region in which the district is located.	Nominal (1) Central (2) Eastern (3) Northern (4) Western
v06	ANC	Total number of 1st antenatal visits by women in the district	Discrete
v07	Season	Changes in weather conditions in the country	Nominal • 1 Dry • 0 Wet
v08	Observ	A time variable measuring the time factor (number of months) from the start (Jan 2015) to the end of the observation period (Dec 2023).	Discrete
v09	IntervenOnset	Measuring time elapsed (number of months) since the intervention time (post SDGs introduction).	Discrete
v10	Interven	A binary variable indicating whether a data point is before or after the intervention (start of SDGs).	Nominal (1) 0 2015 (2) 1 2016 and after

2.3. Mathematical Models. Two models are used in this paper: Bayesian dynamic linear model (BDLM) as explained in the authors' recent article [1] and a mixed effects interrupted time series model (MEITSM) which is this paper's extension of the standard ITS model.

2.3.1. *Bayesian Dynamic Linear Model.* The model was formulated as follows:

Model formulation.

$$Y_{it} = \beta_{0t} + \sum_{i=1}^{n}\sum_{m=1}^{nm}\sum_{t=1}^{h}\beta_{imt}x_{imt} + \sum_{i=1}^{n}\sum_{t=1}^{h}\sum_{m=1}^{nm}\beta_{im(m+1)t}x_{imt}x_{im(m+1)t} + \varepsilon_t \tag{1}$$

$$\begin{bmatrix} \beta_{0t} \\ \beta_{11t} \\ \beta_{112t} \\ \vdots \\ \beta_{nmh} \\ \beta_{nm(m+1)h} \end{bmatrix} = \begin{bmatrix} \beta_{0(t-1)} \\ \beta_{1m(t-1)} \\ \beta_{112(t-1)} \\ \vdots \\ \beta_{nm(h-1)} \\ \beta_{nm(m+1)(h-1)} \end{bmatrix} + \omega_t$$

where:

Y_{it} represents the i^{th} district neonatal mortality rate at time t.

$x_{11}, x_{12}, \ldots, x_{nh}$ are the m covariates including the maternal healthcare policies and interventions in which case a code 0 is used for Before and 1 for After the introduction of a given policy/intervention during the observation time for a given policy or intervention.

$x_{itm}x_{itm(m+1)}$ is the interaction covariate between the m^{th} *and* $(m+1)^{th}$ covariate including the maternal healthcare policies and interventions with $\beta_{itm(m-1)}$ as the their respective coefficients, β_{0t} denotes the time varying intercept.

$\beta_{111}, \beta_{121}, \ldots, \beta_{nhm}$ are the time varying coefficient associated with each of the m covariates including the maternal healthcare policies and interventions.

ε_t and ω_t are the observation and evolution error terms respectively where $\varepsilon_t \sim \mathcal{N}(0, V_t)$ and $\omega_t \sim \mathcal{N}(0, W_t)$.

2.3.2. *Mixed effects interrupted time series model.* The model was formulated as follows:

Study Model formulation.

$$Y_{ijt} = \beta_0 + \beta_1 \cdot \text{Observ}_{ijt} + \beta_2 \cdot \text{Interven}_{ijt} + \beta_3 \cdot \text{IntervenOnset}_{ijt} + \beta_4 \cdot \text{ANC}_{ijt} + \beta_5 \cdot \text{Season}_{ijt} + \beta_6 \cdot \text{Region}_{ijt} + u_i + v_j + \varepsilon_{ijt} \tag{2}$$

where:

Y_{ijt} : Neonatal deaths in region j, time t, for observation i

β_0 : Overall intercept

$\beta_1, \ldots, \beta_6$: Fixed effect coefficients for the predictors

$u_i \sim \mathcal{N}(0, \sigma_u^2)$: Random effect for the spatial variable District

$v_j \sim \mathcal{N}(0, \sigma_v^2)$: Random effect for Region

$\varepsilon_{ijt} \sim \mathcal{N}(0, \sigma^2)$: Residual error term

and :

Observ_{ijt} : Time trend (months since the start of the study)

Interven_{ijt} : Intervention dummy (0 before SDGs, 1 after)

$\text{IntervenOnset}_{ijt}$: Time after intervention began (0 before SDGs, increasing afterward)

ANC_{ijt} : Monthly antenatal care visits

Season_{ijt} : Seasonal effect (dry or wet)

Region_{ijt} : Region fixed effect (Central, Eastern, Northern, Western)

Simplified/general form of the model.

$$y = X\beta + Zu + \varepsilon \tag{3}$$

where:

- $y \in \mathbb{R}^n$ is the response vector (neonatal deaths),
- $X \in \mathbb{R}^{n \times p}$ is the fixed-effects design matrix,
- $\beta \in \mathbb{R}^p$ is the vector of fixed effect coefficients,
- $Z \in \mathbb{R}^{n \times q}$ is the random-effects design matrix.

2.4. Parameter Estimation.

2.4.1. *Parameter Estimation for BDLM.* BDLM parameter estimation was done as follows, as fully explained in the preliminary study by the authors [1] for the formulation of the model. Since the study was interested in the neonatal mortality situation in the country with the introduction of SDGs, it had to make a comparison of the situation before and after their introduction, implying that past values have much influence on future values for this study. The Bayesian approach using the Kalman Filtering technique was considered the best method for estimating study model parameters. Because the Bayesian approach concentrates on showing how a particular part of the data depends on the other parts, it was considered the best approach.

Because Bayesian inference has its roots in Bayes' rule, as noted in the studies by [3, 4]. To get the final parameter estimates, the study made use of the theorem through the posterior density function based on:

$$P(\boldsymbol{\beta}|\boldsymbol{Y}) \propto P(\boldsymbol{Y}|\boldsymbol{\beta})P(\boldsymbol{\beta}) \tag{4}$$

where

$\boldsymbol{\beta}$ represents a vector of parameters to be estimated

Y represents the observed data which is the district's monthly neonatal mortality.

P($\boldsymbol{\beta}$) represents the probability distribution of the parameter vector being estimated, that is to say, the update in the light of new data. This constitutes the prior probability.

$P(\boldsymbol{Y}|\boldsymbol{\beta})$ is the probability distribution of the data given the parameter being estimated which constitutes the likelihood.

$P(\boldsymbol{\beta}|\boldsymbol{Y})$ is the probability distribution of the parameter vector being estimated given the data Y. This constitutes the posterior probability of the estimated parameter (updated upon new data).

Procedure followed to estimate the model parameters

Rewriting the formulated model presented in (1) which is expressed in the form of two equations; observation and state equations, bearing in mind the fact that the process of bayesian inference involves passing from a prior distribution, $P(\boldsymbol{\beta})$ to a posterior distribution, $P(\boldsymbol{\beta}|\boldsymbol{Y})$, estimation of its parameters was done through the Kalman filtering technique.

$$Y_t = \boldsymbol{X}_t^T \boldsymbol{\beta}_t + \varepsilon_t \tag{5}$$

$$\boldsymbol{\beta_t} = \boldsymbol{\beta}_{t-1} + \omega_t$$

where

Y_t is the observed data $\mathbf{X}_t$ is a p-dimensional design vector of covariates an the interaction terms.

$\boldsymbol{\beta}_t$ is a p-dimensional time varying parameter vector (for both covariates an the interaction terms).

ε_t and ω_t are the observation and evolution error terms respectively where $\varepsilon_t \sim \mathcal{N}(0, V_t)$ and $\omega_t \sim \mathcal{N}(0, W_t)$.

Since the formulated model is a multivariate Gaussian state space model, basing on the standard results about a multivariate Gaussian distribution pertaining its marginal and conditional distributions, the random vector of parameters $\boldsymbol{\beta}_t$ too has a Gaussian distribution and the same applies to the other respective marginal and conditional distributions.

Therefore, since all the relevant distributions are Gaussian, we use the means and variances to estimate the parameters (posterior means) through the following procedure:

(1)) Obtaining the prior distribution

From the second line of equation (1), because a DLM is a Gaussian time space model, a normal prior distribution was used for a p-dimensional state vector. First obtaining the initial values where $\hat{\boldsymbol{\beta}}_0 \sim \mathcal{N}(m_0, c_0)$ and then attaining a one step ahead of the prediction of the parameter vector distribution $\hat{\beta}_t | y_{1:t-1}$, we write the prior probability density function as follows:

$$P(\boldsymbol{\beta}_t) = \frac{1}{\sqrt{2\pi(W_t + c_{t-1})^2}} \exp - \frac{(\beta_t - m_{t-1})^2}{2(W_t + c_{t-1})^2} \tag{6}$$

(2)) Obtaining the likelihood distribution function

From the first line of equation (1), obtaining the predicted (one-step) distribution function of Y_t: $Yt|y_{1:t-1}$. Depending on normal distribution conditions, we write its likelihood function probability density function as:

$$P(Y_t|\boldsymbol{\beta}_t) = \frac{1}{\sqrt{2\pi(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2}} \exp - \frac{(Y_t - \boldsymbol{X}_t'\boldsymbol{\beta}_t)^2}{2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2} \tag{7}$$

(3)) Obtaining the posterior distribution

To obtain the posterior distribution, we multiply the likelihood and prior probability density functions as derived from Bayes' theorem in equation (4). Since both the prior and likelihood are Gaussian, the resulting posterior is also Gaussian, represented as $\beta_t|Y_t \sim \mathcal{N}(d_t, p_t)$. This allows us to determine the mean and variance of the parameter estimates by simplifying the posterior pdf expression as illustrated below:

(8)

$$\begin{aligned}
&P(\boldsymbol{\beta}_t|Y_t) \\
&= \frac{1}{\sqrt{2\pi(W_t + c_{t-1})^2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2}} \exp - \frac{-2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2(\beta_t - m_{t-1})^2 - 2(W_t + c_{t-1})^2(Y_t - \boldsymbol{X}_t'\boldsymbol{\beta}_t)^2}{4(W_t + c_{t-1})^2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2} \\
&= \frac{1}{\sqrt{2\pi(W_t + c_{t-1})^2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2}} \exp \frac{-(W_t + c_{t-1})^2((Y_t - \boldsymbol{X}_t'\boldsymbol{\beta}_t)^2)}{2(W_t + c_{t-1})^2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2} \\
&= \frac{1}{\sqrt{2\pi(W_t + c_{t-1})^2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2}} \exp \frac{-(Y_t - \boldsymbol{X}_t'\boldsymbol{\beta}_t)^2}{2(V_t + \boldsymbol{X}_t c_{t|t-1} \boldsymbol{X}_t')^2}
\end{aligned}$$

Therefore, $\boldsymbol{\beta}_t|Y_t \sim \mathcal{N}(\varepsilon_t, (W_t + c_{t-1})^2)$

2.4.2. *Parameter Estimation for MEITSM.* In order to estimate the model parameters, the following aumptions have been put n place basing on equation (3).

Assumptions.

- $u \sim \mathcal{N}(0, G)$ is the vector of random effects,
- $\varepsilon \sim \mathcal{N}(0, R)$ is the residual error vector.
- u and ε are independent.

We marginalize y over u in equation (3) to obtain the marginal distribution of y. We do this by finding the expectation and covariance of y such that:

$E(y) = X\beta$,

$cov(\boldsymbol{y}|\boldsymbol{\beta}) = cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbb{I}_N$

$cov(\boldsymbol{y}) = cov(\boldsymbol{Zu}) + cov(\boldsymbol{\varepsilon}) = \boldsymbol{V}$

therefore:

$$y \sim \mathcal{N}(X\beta, V), \quad \text{where } V = ZGZ^\top + R \quad \text{which is the marginal variance matrix} \tag{9}$$

As a way of getting the log likelihood of REML, we begin by finding the full Maximum Likelihood Estimation of equation (3) we first obtain the the pdf of equation (9) given by:
$p(y) = \int p(y|u) \times p(u)du$
We well know that linear combinations of multivariate normal produce another normal, therefore:
$p(y) = \frac{1}{2\pi|V|^1/2} \exp(\frac{-1}{2}(y - X\beta)^T V^{-1}(y - X\beta))$
Taking the log likelihood, we attain the marginal log-likelihood which is the log-likelihood function under full maximum likelihood as follows:

$$\ell(\theta, \beta) = -\frac{1}{2}\left[n\log(2\pi) + \log|V| + (y - X\beta)^\top V^{-1}(y - X\beta)\right] \tag{10}$$

To obtain $\hat{\beta}$, differentiate $\ell(\theta, \beta)$ with respect to β:

$$\frac{\partial \ell}{\partial \beta} = X^\top V^{-1}(y - X\beta)$$

Setting this derivative to zero gives the Generalized Least Squares (GLS) estimator:

$$\hat{\beta} = (X^\top V^{-1} X)^{-1} X^\top V^{-1} y \tag{11}$$

Substituting $\hat{\beta}$ back into the log-likelihood yields the *profile likelihood*:

$$\ell_P(\theta) = \ell(\theta, \hat{\beta}) = -\frac{1}{2}\left[n\log(2\pi) + \log|V| + (y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta})\right]$$

We then carried out REML estimation to remove the dependence on β by considering only those linear combinations of y that are orthogonal to the column space of X. Let $K \in \mathbb{R}^{n\times(n-p)}$ be a matrix such that:

$$K^\top X = 0 \quad \text{and} \quad K^\top K = I_{n-p}$$

Premultiplying equation (3.3) by $K^\top$ yields:

$$K^\top y \sim \mathcal{N}(0, K^\top V K) \tag{12}$$

The quadratic term in the likelihood becomes:

$$(K^\top y)^\top (K^\top VK)^{-1}(K^\top y) = y^\top K(K^\top VK)^{-1}K^\top y$$

This projects the response vector y into the orthogonal complement of the column space of X, thereby removing fixed effects.

Thus, the REML log-likelihood function is:

$$\log L_{\text{REML}}(\theta) = -\frac{1}{2}\left[(n-p)\log(2\pi) + \log|K^\top VK| + y^\top K(K^\top VK)^{-1}K^\top y\right]$$

Using the following matrix identities:

$$|K^\top VK| = \frac{|V|}{|X^\top V^{-1}X|}, \quad P_\perp = V^{-1} - V^{-1}X(X^\top V^{-1}X)^{-1}X^\top V^{-1}$$

The matrix $P_\perp$ is known as the *generalized orthogonal projection matrix* onto the orthogonal complement of the column space of X with respect to the inner product defined by V^{-1}. This matrix plays a central role in REML because it isolates the portion of the data vector y that is uncorrelated with the fixed effects $X\beta$. Intuitively, $P_\perp$ removes the influence of the fixed effects by projecting the observed data into the subspace orthogonal to X, thereby allowing variance component estimation to proceed without the bias introduced by estimating β.

Hence, the REML log-likelihood becomes:

(13)
$$\log L_{\text{REML}}(\theta) = -\frac{1}{2}\left[(n-p)\log(2\pi) + \log|V| + \log|X^\top V^{-1}X| + (y-X\hat{\beta})^\top V^{-1}(y-X\hat{\beta})\right]$$

To obtain the parameter estimates, we maximized this log-likelihood with respect to θ, the parameters defining G and R. The estimate of β was obtained by the GLS estimator as in expressed in equation (11), which was derived by solving the normal equations:

$$X^\top V^{-1}X\hat{\beta} = X^\top V^{-1}y$$

Thus, from the REML framework, the parameter estimates are:

- $\hat{\beta} = (X^\top V^{-1}X)^{-1}X^\top V^{-1}y$
- $\hat{\theta} = \arg\max_\theta \log L_{\text{REML}}(\theta)$

where the latter ($\hat{\theta}$): the random effects estimates, was obtained via numerical optimization such as the Newton-Raphson method. with this approach we attained consistent and unbiased estimates of both the fixed effects and the variance components (Random efects) in the mixed-effects ITS model.

2.5. Model Assumptions. For the two models above, the following properties were checked for:

Auto-correlation:

The Residuals should not be autocorrelated.

Homoscedasciticity:

The residuals should have constant Variance.

Stationarity:

The residuals are expected to be stationary.

2.6. Model Diagnostics (Robust tests). To validate the assumptions of the two models specified in (2) and (1), we tested the residuals for autocorrelation, heteroscedasticity, and stationarity (unit root).

1. Autocorrelation: Durbin-Watson Test.

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{14}$$

where

e_t = the residuals from the regression model.

T = the number of observations.

Hypotheses:

- H_0: No first-order autocorrelation
- H_1: First-order autocorrelation exists

Decision rule:

- If $DW \approx 2$: No autocorrelation
- If $DW < 2$: Positive autocorrelation
- If $DW > 2$: Negative autocorrelation

Importantly, the p-value of the Durbin Watson test statistic should be greater than 0.05 for us to conclude that the residuals are not autocorrelated.

2. Heteroscedasticity: Breusch-Pagan Test. **Test statistic:**

$$\text{BP} = \frac{SSR^*}{2} \tag{15}$$

where SSR^* is the regression sum of squares from an auxiliary regression of squared residuals $\hat{e}^2$ on the original regressors.

Hypotheses:

- H_0: Homoscedasticity: (Residuals have constant variance)
- H_1: Heteroscedasticity exists (Residuals do not have constant variance)

Decision rule: If the probability value of the Breusch Pagan test statistic is greater than or equal to 0.05 ($p-value \geq 0.05$), we fail to reject the null hypothesis and conclude that Residuals have constant variance. Otherwise, we reject the null hypothesis and conclude that the Residuals do not have constant variance.

3. Unit Root: Augmented Dickey-Fuller (ADF) Test.

$$\Delta e_t = \alpha + \gamma e_{t-1} + \sum_{i=1}^{p} \delta_i \Delta e_{t-i} + \varepsilon_t \tag{16}$$

Test statistic:

$$\tau = \frac{\hat{\gamma}}{\text{SE}(\hat{\gamma})} \tag{17}$$

where

$\hat{\gamma}$ = is the estimated coefficient of Y_{it-1}, and

S.E($\hat{\gamma}$) its standard error.

Hypotheses:

- H_0: Residuals have a unit root (non-stationary)
- H_1: Residuals are stationary

Decision rule:

Reject H_0 if τ is less than the critical value at a chosen significance level. For instance, at 5% significance level, if the probability value of the ADF test statistic is greater than or equal

to 0.05 ($p - value \geq 0.05$), we fail to reject the null hypothesis and conclude that the data is non-stationary.

2.7. Model Evaluation using Mean Cross-Validated Log-Likelihood. Based on the literature by [7, 13, 6], we were able to come up with a procedure to assess the out-of-sample predictive performance of the two models. We employed the Mean Cross-Validated Log-Likelihood (MCVLL) and Cross-Validated Root Mean Squared Error (CVRMSE) using K-fold cross-validation technique.

Let $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^{n}$ be a dataset of size n, where x_i represents the covariates and y_i the response variable (monthly neonatal deaths). In K-fold cross-validation, the dataset is randomly partitioned into K approximately equal-sized, mutually exclusive subsets (or folds) denoted by $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_K$ ($K = 3$).

For each fold $k \in \{1, 2, \ldots, K\}$:

- The test set is defined as $\mathscr{D}_k$.
- The training set is defined as $\mathscr{D}_{-k} = \mathscr{D} \setminus \mathscr{D}_k$.
- The model fitted on the training set is denoted by $\hat{f}^{(k)}$.

2.7.1. *Mean Cross-Validated Log-Likelihood.* The mean cross-validated log-likelihood (MCVLL) assesses how well a model predicts unseen data by averaging the log-predictive densities across all test folds:

$$\text{MCVLL} = \frac{1}{n} \sum_{k=1}^{K} \sum_{(x_i, y_i) \in D_k} \log p(y_i \mid x_i, \hat{f}^{(k)}) \tag{18}$$

The predictive distribution is given by:

$$p(y_i \mid x_i, D_{-k}) = \int p(y_i \mid x_i, \theta)\, p(\theta \mid D_{-k})\, d\theta \tag{19}$$

Since this integral is intractable, it is approximated using S posterior samples $\{\theta^{(s)}\}_{s=1}^{S}$:

$$p(y_i \mid x_i, D_{-k}) \approx \frac{1}{S} \sum_{s=1}^{S} p(y_i \mid x_i, \theta^{(s)}) \tag{20}$$

The log-predictive density is then estimated using:

$$\log p(y_i \mid x_i, D_{-k}) \approx \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i \mid x_i, \theta^{(s)}) \right) \tag{21}$$

To ensure numerical stability, especially when dealing with small likelihood values, the log-sum-exp trick is used:

$$\log\left(\frac{1}{S}\sum_{s=1}^{S}\exp(\ell_s)\right) = -\log S + \text{log-sum-exp}(\ell_1,\ldots,\ell_S) \tag{22}$$

where $\ell_s = \log p(y_i \mid x_i, \theta^{(s)})$.

2.7.2. *Cross-Validated Root Mean Squared Error (CVRMSE).* The CV-RMSE evaluates predictive performance in terms of squared error and is computed as:

$$\text{CVRMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{K}\sum_{(x_i,y_i)\in D_k}\left(y_i - \hat{y}_i^{(k)}\right)^2} \tag{23}$$

where $\hat{y}_i^{(k)}$ is the prediction for x_i from the model fitted on D_{-k}. RMSE provides an interpretable error metric in the same units as the response variable, making it useful for comparing competing models [7].

2.7.3. *Model Comparison Strategy (Decision Rule).* To compare two candidate models, we use both metrics:

- A model with a higher MCVLL is preferred for its superior predictive likelihood.
- A model with a lower CVRMSE is preferred for its smaller average prediction error.

These cross-validated criteria were considered since they provide a principled and data-driven approach to model selection in both frequentist and Bayesian contexts.

3. RESULTS

3.1. Model validation.

3.1.1. *Robust methods of model Validation.*

TABLE 2. Empirical mean and its standard error for each variable

Variable	Mean	std.error
ANC	0.006	0.000
After Interven	-0.332	0.018
Before Interven	1.664	0.021
IntervenOnset	-0.181	0.022
Observ	0.170	0.021
RegionCentral	1.099	0.017
RegionEastern	0.338	0.018
RegionNorthern	1.051	0.022
RegionWestern	0.463	0.019
RegionCentralEastern	0.901	0.002
RegionCentralNorthen	0.899	0.001
RegionCentralWestern	0.901	0.001
RegionEasternNorthern	0.901	0.002
RegionEasternWestern	0.899	0.002
RegionNorthernWestern	0.901	0.001
SeasonDry	0.583	0.013
SeasonWet	0.779	0.014

TABLE 3. Regression output showing the parameter estimate with its standard error from the Mixed Effects ITS Model

Variable	Estimate	std.error
(Intercept)	6.730	4.517
ANC	0.001	0.000
After Interven	-5.334	0.383
IntervenOnset	-1.618	0.055
Observ	0.605	0.055
RegionEastern	-5.424	6.283
RegionNorthern	-4.861	6.281
RegionWestern	-4.590	6.290
SeasonWet	0.137	0.125

TABLE 4. Models validation tests Results

Model/Test	DW Test for Autocorrelation		BP Test for Heteroscedasticity		ADF Test for Stationarity	
	Statistic	**p-value**	**Statistic**	**p-value**	**Statistic**	**p-value**
MEITSM	1.4385	0.00	799.87	2.2e-16	-13.027	0.01
BDLM	0.5864	0.00	0.25	0.6	-13	0.01

3.1.2. *Mean Log-Likelihood method of K-Fold Cross-Validation.*

TABLE 5. Comparison of Mean Log-Likelihoods from K-Fold Cross-Validation

Model	**Mean CV Log-Likelihood**	**Interpretation**
Interrupted Time Series (MEITSM)	-3.7631	Poorer predictive performance
Dynamic Linear Model (BDLM)	-3.3025	Better predictive performance

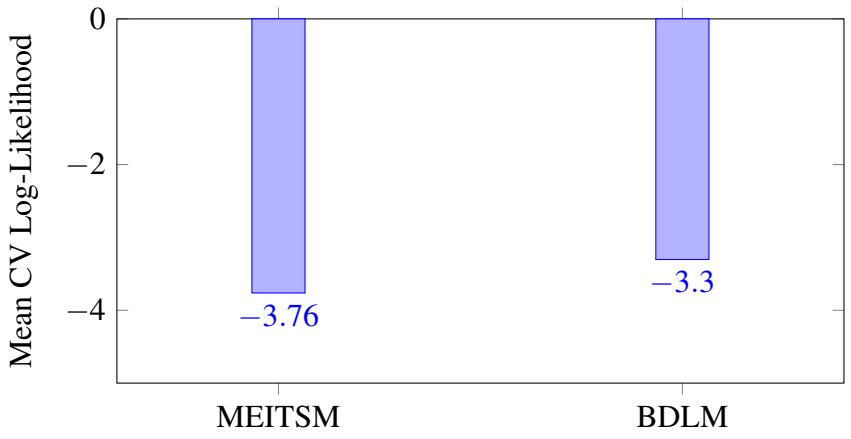

FIGURE 1. Bar plot of mean log-likelihood from K-fold cross-validation for MEITSM and BDLM

3.1.3. *RMSE method of K-Fold Cross-Validation comparison.*

TABLE 6. Comparison of RMSE from K-Fold Cross-Validation

Model	Mean CV RMSE	Interpretation
Mixed Effects Interrupted Time Series Model (MEITSM)	7.662	Better predictive performance
Bayesian Dynamic Linear Model (BDLM)	15.712	Substantial predictive performance

4. DISCUSSION OF RESULTS

Basing on the results in output in Table 4:

While testing for autocorrelation of residuals ,since the p-value of the Durbin-Watson statistic is 0.00 for both models, this indicates autocorrelation of the residuals, hence both models fail on this property.

While testing for homoscedasticity of residuals, by the fact that the probability value of the of Breusch Pagan test statistic for the BDLM (p-value = 0.6) is greater than 0.05, we fail to reject the null hypothesis and conclude that Residuals do have constant variance while Residuals do not have constant variance for MEITSM since its Breusch Pagan test statistic p-value (2.2e-16) is less than 0.05. Therefore, the MEITS does not fulfill a very important property of homoscedasticity of residuals.

While testing whether the residuals are stationary, whereby the Augmented Dickey-Fuller test was used, since the p-value of the ADF test statistic is less than 0.05 $(p-value \geq 0.05)$ for both models, we reject the null hypothesis and conclude that the two models have stationary residuals.

Therefore, since MEITSM fails on two occasions, especially on homoscedasticity, we cannot rely on it and so we take the BDLM since it only fails on autocorrelation, a property that can be overlooked with non-frequentist models.

Comparing regression output in Table 2 which is obtained using the formulated study model (BDLM) with that in Table 3 which is obtained using the Mixed Effects ITS model (MEITSM), there is evidence based on the standard errors that the Formulated study model (BDLM) provides better estimates, as it has lower standard errors.

In addition, on top of the previous results in Table 4 from model validation tests on autocorrelation, heteroskedasticity and unit root (stationarity) of residuals, based on the Global measure of goodness of fit, to assess and compare the predictive performance of the mixed effects Interrupted Time Series model (MEITSM) and the Bayesian Dynamic Linear Model (BDLM), we applied K-fold cross-validation to both models. The mean log-likelihood across folds was used as the performance metric, where a higher (less negative) value indicates better predictive accuracy.

From Tables 5 and 6, Although the Mixed Effects Interrupted Time Series Model (MEITSM) demonstrated a lower root mean square error (RMSE) during cross-validation, indicating superior point-wise predictive accuracy, the Bayesian Dynamic Linear Model (BDLM) exhibited a higher mean cross-validation log-likelihood as further shown in Figure 1, suggesting better probabilistic modeling of the observed data distribution.

5. CONCLUSION

Due to the fact that BDLM flexibly captures temporal dynamics and structural changes in neonatal mortality, it offers a more robust and comprehensive framework for evaluating the evolving impact of the Sustainable Development Goals (SDGs) onset over time. Furthermore, the BDLM's Bayesian formulation provides credible intervals and probabilistic forecasts that enhance interpretability and decision-making in policy contexts. Therefore, despite the higher RMSE, BDLM outshines MEITSM as a preferred model for Neonatal mortality analysis amidst SDGs onset impact evolution, in alignment with the study's methodological innovations and overarching research objectives.

ACKNOWLEDGMENTS

The authors acknowledge the African Union for sponsoring the study.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] B. George, G.A. Waititu, R.O. Awichi, A.O. Adebanji, Formulation of a Spatiotemporal Model for Analysis of Neonatal Mortality Amidst Sdgs Intervention. the Case of Uganda, medRxiv (2025). https://doi.org/10.1101/2025.04.17.25326005.

[2] J. Lopez Bernal, S. Cummins, A. Gasparrini, Interrupted Time Series Regression for the Evaluation of Public Health Interventions: A Tutorial, Int. J. Epidemiol. 46 (2017), 348–355. https://doi.org/10.1093/ije/dyw098.

[3] C.M. Bishop, N.M. Nasrabadi, Pattern Recognition and Machine Learning, Springer, 2006.

[4] A. Eshky, Bayesian Methods of Parameter Estimation, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/eshky.pdf.

[5] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013. https://doi.org/10.1201/9780429258411.

[6] A. Gelman, J. Hwang, A. Vehtari, Understanding Predictive Information Criteria for Bayesian Models, Stat. Comput. 24 (2014), 997–1016. https://doi.org/10.1007/s11222-013-9416-2.

[7] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[8] E. Kontopantelis, T. Doran, D.A. Springate, I. Buchan, D. Reeves, Regression Based Quasi-Experimental Approach When Randomisation Is Not an Option: Interrupted Time Series Analysis, BMJ 350 (2015), h2750. https://doi.org/10.1136/bmj.h2750.

[9] Ministry of Health Uganda, Annual Health Sector Performance Report 2019/2020, 2020. https://www.health.go.ug/wp-content/uploads/2020/11/Annual-Health-Sector-Performance-Report-2019-2020-FY-1.pdf.

[10] United Nations, Transforming Our World: The 2030 Agenda for Sustainable Development, 2015. https://sdgs.un.org/2030agenda.

[11] WHO, The Health of Mothers and Babies Is the Foundation of Healthy Families and Communities, World Health Organization, 2025. https://www.afro.who.int/countries/uganda/news/health-mothers-and-babies-foundation-healthy-families-and-communities.

[12] UNICEF, Maternal and Newborn Health Disparities: Uganda, 2021. https://data.unicef.org/wp-content/uploads/cp/maternal-newborn-health/maternal-and-newborn-health_UGA.pdf.

[13] A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC, Stat. Comput. 27 (2016), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4.

[14] A.K. Wagner, S.B. Soumerai, F. Zhang, D. Ross-Degnan, Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research, J. Clin. Pharm. Ther. 27 (2002), 299–309. https://doi.org/10.1046/j.1365-2710.2002.00430.x.