BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 1229–1236
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp1229-1236

THE UNINFORMATIVE PRIOR OF JEFFREYS' DISTRIBUTION IN BAYESIAN GEOGRAPHICALLY WEIGHTED REGRESSION

Fachri Faisal [1*], Henny Pramoedyo[2], Suci Astutik[3], Achmad Efendi[4]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya*
[2,3,4]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya*
Jln. Veteran, Malang, 65145, Indonesia

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Bengkulu*
Jl. W.R. Supratman, Kandang Limun, Bengkulu, 38371, Indonesia

*Corresponding author's e-mail: *fachrif@unib.ac.id*

ABSTRACT

Article History:

Received: 9th January 2024
Revised: 15th February 2024
Accepted: 15th April 2024
Published: 1st June 2024

Keywords:

Marginal Posterior Distributions;
Prior And Posterior;
Uninformative Jeffreys Prior.

When using the Bayesian method for estimating parameters in a geographically weighted regression model, the choice of the prior distribution directly impacts the posterior distribution. The distribution known as the Jeffreys prior is an uninformative type of prior distribution and is invariant to reparameterization. In cases where information about the parameter is not available, the Jeffreys' prior is utilized. The data was fitted with an uninformative Jeffreys' prior distribution, which yielded a posterior distribution that was utilized for estimating parameters. This study aims to derive the prior and marginal posterior distributions of the Jeffreys' $\boldsymbol{\beta}(S)$ *and* $\sigma^2(S)$ *in Bayesian geographically weighted regression (BGWR). The marginal posterior distributions of* $\boldsymbol{\beta}(S)$ *and* $\sigma^2(S)$ *can be obtained by integrating the other parameters of a common posterior distribution. Based on the results and discussion, the Jeffreys prior in BGWR with the likelihood function* $\boldsymbol{Y}|\boldsymbol{\beta}(s),\boldsymbol{X},\boldsymbol{W}(s),\sigma^2(s) \sim MVN\left(\boldsymbol{X\beta}(s),\sigma^2(s)\boldsymbol{W}^{-1}(s)\right)$ *is* $|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}\left(\sigma^2(s)\right)^{-(p+3)/2}$*. On the other hand, the marginal posterior distribution of* $\boldsymbol{\beta}(S)$ *follows a normal multivariate distribution, that is,* $\boldsymbol{\beta}(S) \sim MVN\left(\hat{\boldsymbol{\beta}}(s),\sigma^2(s)(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X})^{-1}\right)$*, while the marginal posterior distribution of* $\sigma^2(s)$ *follows an inverse gamma distribution, that is,* $\sigma^2(s) \sim IG\left(\frac{n+1}{2},\frac{\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{HY}}{2}\right)$*. As further research, it is necessary to follow up on several limitations of the results of this research, namely numerical simulations and application to a particular case that related to the results of the analytical studies that we have carried out.*

How to cite this article:

F. Faisal, H. Pramoedyo, S. Astutik, A. Efendi., "THE UNINFORMATIVE PRIOR OF JEFFREYS' DISTRIBUTION IN BAYESIAN GEOGRAPHICALLY WEIGHTED REGRESSION," *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 1229-1236, June, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Bayes' theorem was discovered by Thomas Bayes, a cleric, statistician, and philosopher. In 1963, Richard Price, a Minister, philosopher, and mathematician, established the conceptual foundation of Bayesian statistics in the Royal Society. Sir Harold Jeffreys also wrote on Bayes' theorem in a book published in 1973. Thomas Bayes' ideas established the foundation for the notion of Bayesian analysis in estimating unknown model parameters, which developed rapidly and is today extensively employed in a variety of disciplines of research. The Bayesian analysis concept method is extensively used in a variety of industries since it offers several benefits over non-Bayesian approaches (classical / frequentist statistical approaches) for estimating unknown model parameters [1]. It should be noted that this theorem seeks to blend information from data modeled with information from parameters gained from previous occurrences (priors) in order to improve the validity of the model parameters to be estimated. According to this, the optimal model with the highest posterior probability may be chosen, and a conclusion may be drawn regarding the average of the calculated model parameters [2].

The effect of the data on the posterior distribution often increases with the number of observations. Stated differently, the posterior distribution will be more impacted by the data than by the previous distribution as the number of observations rises. On the other hand, the posterior distribution is mostly determined by the prior distribution when there is a dearth of data. The posterior distribution that is created will be mismatched if there is an error in the prior distribution of the parameters, which will lead to the wrong (invalid) estimation of the unknown parameters that should be known. Thus, in order to identify the most suitable and reliable parameter estimator, establishing the prior distribution for the unknown parameter is a crucial and essential component of the Bayesian method estimation idea [2].

In connection with what has been described previously, on this occasion, the author uses the Bayesian Theorem or Bayesian method in the Geographically Weighted Regression (GWR) model. The GWR model can be succinctly described as a localized approach to spatial analysis, where weighting is determined by the proximity or distance between locations of observation. Regression parameters in the GWR model are assumed to vary spatially, so that different and valuable interpretations can be obtained for each location point under study. This approach is based on spatial non-stationarity, which provides flexible parameter estimates, and any spatially nonstationary relationship cannot be represented by global statistics [3]. The application of this method has been widely used across various fields, such as climatology [4], geology [5], criminology [6], transportation analysis [7], house price modeling [8], and forestry [9]. However, the GWR model also has some limitations, namely, the potential for inappropriate distribution of model coefficient estimates, susceptibility to the influence of outliers, challenges related to weak data, and violations of the assumption of homogeneity of error variances.

To overcome these limitations, LeSage, in his article, proposed a Bayesian approach called Bayesian GWR (BGWR), which aims to estimate local coefficients [10]. The BGWR model incorporates the concept of parameter smoothing, which relates the regression of local coefficients from one location to those from other locations in the study area. In addition, this Bayesian approach provides more accurate results in terms of GWR parameter estimation [11]. Research related to the BGWR model can be seen in [9]-[12] using improper prior distribution based on [13], while others used conjugate priors [14]. Based on the results of the study conducted by these authors, it is concluded that the BGWR model is better than the GWR model. One of the factors supporting these results is that the Bayesian approach can directly identify and weight observations that may contain outliers, thereby reducing the impact of outliers on model parameter estimates.

The prior distribution in the context of BGWR is a probability distribution that describes the knowledge of the researcher or confidence in the model parameters. The prior distribution is combined with the likelihood function, which represents the probabilities of the observed data with the parameters present, to obtain the posterior distribution, which represents the updated probabilities of the parameters present with the information of prior and data. After that, the posterior distribution is utilized to make inferences about the parameters and generate predictions for new data [15].

Improper, conjugate, and Jeffreys priors have been widely used in linear regression analysis, while the use in GWR has so far only used improper and conjugate priors. Each type of prior has its own advantages and disadvantages, depending on the research question and the data being analyzed. In general, some disadvantages of improper priors are the problem of lack of information and sensitivity to data while the disadvantages of conjugate priors are limited flexibility and applicability. Jeffreys priors are favored for their objectivity and invariance properties, making them a choice when there is limited knowledge about the data or model and the primary aim is to prevent bias from affecting the model [16], [17], [18]. Furthermore, the Jeffreys prior was first published in 1946, and the main attraction of the Jeffreys prior is that it is invariant under coordinate transformation [19].

Based on the description above, this study emphasizes the process of obtaining Jeffrey's prior distribution through the likelihood function based on [14], which then also obtained the posterior distribution and marginal posterior distribution and finally used for parameter estimation of the Bayesian GWR model. According to these motivations, the research gap that will be studied in this paper is the process of obtaining the Jeffrey prior distribution through studying the likelihood function.

The rest of the paper is organized as follows: The research methods related to this work, such as the BGWR approach, including steps in estimating the parameters and how to obtain Jeffreys' uninformative prior and posterior in the BGWR model, are provided in Section 2. In Section 3, we provide results and discussion related to our work. Finally, some conclusions and recommendations for further study are discussed in Section 4.

2. RESEARCH METHODS

This study uses the BGWR approach, where the first thing to determine is the prior before using the Bayesian model. Next, a brief theory of GWR and Bayesian is presented as well as the research steps.

The GWR model can be expressed as follows from [14]:

$$\boldsymbol{y}(\boldsymbol{s}) = \boldsymbol{\beta}_{\mathbf{1}}(\boldsymbol{s})\boldsymbol{x}_{\mathbf{1}}(\boldsymbol{s}) + \cdots + \boldsymbol{\beta}_{\boldsymbol{p}}(\boldsymbol{s})\boldsymbol{x}_{\boldsymbol{p}}(\boldsymbol{s}) + \boldsymbol{\epsilon}(\boldsymbol{s}), \tag{1}$$

where $\boldsymbol{y}(\boldsymbol{s})$ denotes the dependent variable at location $\boldsymbol{s}$, $\boldsymbol{\beta}_{\boldsymbol{i}}(\boldsymbol{s}), \boldsymbol{i} = \mathbf{1}, \cdots, \boldsymbol{p}$ are the coefficients of explainable variables at location $\boldsymbol{s}$, and the random effect at location $\boldsymbol{s}$, denoted by $\boldsymbol{\epsilon}(\boldsymbol{s})$ is assumed to follow $\boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^{\mathbf{2}})$. To calculate the weights of each observation, one can utilize a weighting function and determine the distance between every observation and $\boldsymbol{s}$.The weighted least squares method can be used to estimate the coefficients at point $\boldsymbol{s}$:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{s}) = \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y}, \tag{2}$$

where $\boldsymbol{X}$ represents the covariate matrix with size $\boldsymbol{n} \times \boldsymbol{p}$, $\boldsymbol{Y}$ is the vector of responses with dimensions $\boldsymbol{n} \times \mathbf{1}$, and $\boldsymbol{W}(\boldsymbol{s})$ is a diagonal matrix, denoted by $\boldsymbol{diag}\left(\boldsymbol{w}_{\mathbf{1}}(\boldsymbol{s}), \cdots, \boldsymbol{w}_{\boldsymbol{n}}(\boldsymbol{s})\right)$**,** which contains the weights,

$$\boldsymbol{W}(\boldsymbol{s}) = \begin{bmatrix} \boldsymbol{w}_{\mathbf{1}}(\boldsymbol{s}) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{w}_{\boldsymbol{n}}(\boldsymbol{s}) \end{bmatrix} \tag{3}$$

Equation (2) estimates the parameters of the GWR model by first adding a weighting factor to **Equation (1)** and then minimizing the weighted sum of squares error.

Furthermore, **Equation (2)** may be expressed as a matrix along with its algebraic operations, as can be seen in **Equation (4)** below.

$$\boldsymbol{\epsilon}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{\epsilon} = \boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y} - \mathbf{2}\boldsymbol{\beta}^{\boldsymbol{T}}(\boldsymbol{s})\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y} + \boldsymbol{\beta}^{\boldsymbol{T}}(\boldsymbol{s})\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}). \tag{4}$$

To estimate the parameter of $\widehat{\boldsymbol{\beta}}(s)$, use the **Equation (4)** as follows:

$$\frac{\partial \boldsymbol{\epsilon}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{\epsilon}}{\partial \boldsymbol{\beta}(\boldsymbol{s})} = -\mathbf{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y} + \mathbf{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}). \tag{5}$$

By **Equation (5)** with zero, the parameter estimate for the GWR model is obtained:

$$\begin{gathered} \mathbf{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}) = \mathbf{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y} \\ \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}) = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y} \\ \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}) = \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y} \end{gathered} \tag{6}$$

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{s}) = \left(\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Y}$$

In this study, we employ Bayesian methods for GWR, utilizing a likelihood-based approach as described in [14]. This model's likelihood function has the following form:

$$\boldsymbol{Y}|\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{X},\boldsymbol{W}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s}) \sim \boldsymbol{MVN}\left(\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})\boldsymbol{W}^{-1}(\boldsymbol{s})\right), \tag{7}$$

where the likelihood function has a multivariate normal distribution (MVN).

The posterior distribution for which inference is

$$\boldsymbol{f}\left(\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})|\boldsymbol{Y}\right) = \frac{\boldsymbol{f}\left(\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})\right)\boldsymbol{f}\left(\boldsymbol{Y}|\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})\right)}{\boldsymbol{f}(\boldsymbol{Y})}, \tag{8}$$

where $f(\boldsymbol{Y})$ is the marginal distribution of the data and in general, $f(\boldsymbol{Y})$ contains no parameters and is a constant value, then **Equation (8)** can be expressed in the equation

$$\boldsymbol{f}\left(\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})|\boldsymbol{Y}\right) \propto \boldsymbol{f}\left(\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})\right)\boldsymbol{f}\left(\boldsymbol{Y}|\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})\right). \tag{9}$$

According to **Equation (5)**, the posterior is directly proportional to the product of the prior of the model parameters and the likelihood [20], [21].

To obtain the Jeffreys' uninformative prior and posterior in the Bayesian GWR model, the following process is carried out:

i. Write down the likelihood function. In this case, the likelihood function is in **Equation (7)**
ii. Compute the Fisher information matrix, which is formulated by $I\left(\boldsymbol{\beta}(s),\sigma^2(s)\right) = -E\left[\frac{\partial^2 \log f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s),\boldsymbol{X},\boldsymbol{W}(s),\sigma^2(s)\right)}{\partial\left(\boldsymbol{\beta}(s),\sigma^2(s)\right)^2}\right]$.
iii. Calculate $\left|I\left(\boldsymbol{\beta}(s),\sigma^2(s)\right)\right|$.
iv. Calculate the Jeffreys prior, which is expressed by $f\left(\boldsymbol{\beta}(s),\sigma^2(s)\right) \propto \left|I\left(\boldsymbol{\beta}(s),\sigma^2(s)\right)\right|^{1/2}$
v. Get the posterior distribution by multiplying the Jeffreys prior by the likelihood function.
vi. Obtained from (v), the marginal posterior distribution of each parameter.

3. RESULTS AND DISCUSSION

1. The likelihood functions

The likelihood function in **Equation (7)** is first determined as follows:

$$f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s),\boldsymbol{X},\boldsymbol{W}(s),\sigma^2(s)\right) = \left(\boldsymbol{2\pi\sigma^2(s)}\right)^{-\boldsymbol{n/2}}|\boldsymbol{W}(\boldsymbol{s})|^{\boldsymbol{1/2}}\mathbf{exp}\left\{-\frac{\boldsymbol{1}}{\boldsymbol{2\sigma^2(s)}}\left(\boldsymbol{Y}-\boldsymbol{X\beta}(\boldsymbol{s})\right)^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{s})\left(\boldsymbol{Y}-\boldsymbol{X\beta}(\boldsymbol{s})\right)\right\} \tag{10}$$

2. The uninformative Jeffreys prior distribution

To calculate the Jeffreys' prior in Bayesian GWR with the likelihood function in **Equation (7)**, we need to follow these steps:

i. Derive the first partial derivative of $\boldsymbol{\beta}$ simultaneously using matrix calculus:

$$\frac{\boldsymbol{\partial}\mathbf{log}\,\boldsymbol{f}\left(\boldsymbol{Y}|\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{X},\boldsymbol{W}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})\right)}{\boldsymbol{\partial\beta}(\boldsymbol{s})} = \frac{\partial}{\partial\boldsymbol{\beta}(s)}\left(-\frac{1}{2\sigma^2(s)}\left(\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y} - 2\boldsymbol{\beta}^T(s)\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right)\right)$$

$$= \frac{1}{2\sigma^2(s)} \frac{\partial}{\partial \boldsymbol{\beta}(s)} \left(2\boldsymbol{\beta}^T(s)\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{Y} - \boldsymbol{\beta}^T(s)\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right)$$
$$= \frac{1}{2\sigma^2(s)}\left(2\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{Y} - 2\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right)$$
$$= \frac{1}{\sigma^2(s)}\left(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right). \tag{11}$$

Now the second partial derivative:

$$\frac{f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s), \boldsymbol{X}, \boldsymbol{W}(s), \sigma^2(s)\right)}{\partial \boldsymbol{\beta}^2(s)} = \frac{\partial^2}{\partial \boldsymbol{\beta}^2(s)}\left(\frac{1}{\sigma^2}\left(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right)\right)$$
$$= \frac{1}{\sigma^2(s)} \frac{\partial^2}{\partial \boldsymbol{\beta}^2(s)}\left(-\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right)$$
$$= -\frac{1}{\sigma^2(s)}(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}) \tag{12}$$

Now taking the negative expectation:

$$-E\left[\frac{\partial^2 f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s), \boldsymbol{X}, \boldsymbol{W}(s), \sigma^2(s)\right)}{\partial \boldsymbol{\beta}^2(s)}\right] = \frac{1}{\sigma^2(s)}(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}). \tag{13}$$

ii. Derive the first partial derivative of $\sigma^2(s)$ simultaneously:

$$\frac{\partial \log f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s), \boldsymbol{X}, \boldsymbol{W}(s), \sigma^2(s)\right)}{\partial \sigma^2(s)}$$
$$= \frac{\partial}{\partial \sigma^2(s)}\left(-\frac{n}{2}\log\left(\sigma^2(s)\right) - \frac{1}{2\sigma^2(s)}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)^T\boldsymbol{W}(s)\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)\right)$$
$$= -\frac{n}{2\sigma^2(s)} + \frac{1}{2\left(\sigma^2(s)\right)^2}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)^T\boldsymbol{W}(s)\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right). \tag{14}$$

Now the second partial derivative:

$$\frac{\partial^2 \log f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s), \boldsymbol{X}, \boldsymbol{W}(s), \sigma^2(s)\right)}{\partial(\sigma^2(s))^2}$$
$$= \frac{\partial^2}{\partial(\sigma^2(s))^2}\left(-\frac{n}{2\sigma^2(s)} + \frac{1}{2\left(\sigma^2(s)\right)^2}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)^T\boldsymbol{W}(s)\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)\right)$$
$$= \frac{n}{2\left(\sigma^2(s)\right)^2} - \frac{1}{\left(\sigma^2(s)\right)^3}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)^T\boldsymbol{W}(s)\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)$$
$$= \frac{n\sigma^2(s) - 2\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)^T\boldsymbol{W}(s)\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(s)\right)}{2\left(\sigma^2(s)\right)^3}. \tag{15}$$

Now taking the negative expectation:

$$-E\left[\frac{\partial^2 f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s), \boldsymbol{X}, \boldsymbol{W}(s), \sigma^2(s)\right)}{\partial(\sigma^2(s))^2}\right] = \left(\frac{-n\sigma^2(s) + 2n\sigma^2(s)}{2\left(\sigma^2(s)\right)^3}\right) = \frac{n}{2\left(\sigma^2(s)\right)^2}. \tag{16}$$

iii. Derive cross products for $\boldsymbol{\beta}(s)$ and $\sigma^2(s)$:

$$-E\left[\frac{\partial^2 f\left(\boldsymbol{Y}|\boldsymbol{\beta}(s), \boldsymbol{X}, \boldsymbol{W}(s), \sigma^2(s)\right)}{\partial \boldsymbol{\beta}(s)\partial \sigma^2(s)}\right] = -E\left[-\frac{2}{\sigma^3(s)}\left(\boldsymbol{X}^T W(s)\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\boldsymbol{\beta}(s)\right)\right]$$
$$= \frac{2}{\sigma^3(s)}\boldsymbol{X}^T\boldsymbol{W}(s)\left(\boldsymbol{X}\boldsymbol{\beta}(s) - \boldsymbol{X}\boldsymbol{\beta}(s)\right) = 0. \tag{17}$$

iv. The Fisher information matrix is:

$$I(\boldsymbol{\beta}(s),\sigma^2(s)) = -E\left[\frac{\partial^2 \log f(\boldsymbol{Y}|\boldsymbol{\beta}(s),\boldsymbol{X},\boldsymbol{W}(s),\sigma^2(s))}{\partial(\boldsymbol{\beta}(s),\sigma^2(s))^2}\right], \tag{18}$$

where $f(\boldsymbol{Y}|\boldsymbol{\beta}(s),\boldsymbol{X},\boldsymbol{W}(s),\sigma^2(s))$ is the likelihood function. For the given likelihood function, the Fisher information matrix is:

$$I(\boldsymbol{\beta}(s),\sigma^2(s)) = \begin{bmatrix} \frac{\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}}{\sigma^2(s)} & 0 \\ 0 & \frac{n}{2(\sigma^2(s))^2} \end{bmatrix}, \tag{19}$$

where n is the number of observations.

v. Let p is the number of parameters. The determinant of the Fisher information matrix is:

$$\begin{aligned} |I(\boldsymbol{\beta}(s),\sigma^2(s))| &= (\sigma^2(s))^{-(p+1)}|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|\frac{n}{2}(\sigma^2(s))^{-2} \\ &= \frac{n}{2}|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|(\sigma^2(s))^{-(p+3)}, \end{aligned} \tag{20}$$

Taking the square root of the determinant in **Equation (20)**, we have:

$$|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}(\sigma^2(s))^{-(p+3)/2}. \tag{21}$$

vi. The Jeffreys prior is then given by:

$$\begin{aligned} f(\boldsymbol{\beta}(\boldsymbol{s}),\boldsymbol{\sigma}^2(\boldsymbol{s})) &\propto |\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}|^{1/2}(\boldsymbol{\sigma}^2(\boldsymbol{s}))^{-(\boldsymbol{p}+3)/2} \\ &\propto |\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}|^{1/2}(\boldsymbol{\sigma}^2(\boldsymbol{s}))^{-\left(\frac{\boldsymbol{p}+1}{2}+1\right)}. \end{aligned} \tag{22}$$

3. The posterior distribution from Jeffreys prior

By using **Equation (9)**, **Equation (10)** and **Equation (22)**, the posterior distribution can be written as:

$$\begin{aligned} &f(\boldsymbol{\beta}(s),\sigma^2(s)|\boldsymbol{Y}) \\ &\propto |\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{s})\boldsymbol{X}|^{1/2}(\boldsymbol{\sigma}^2(\boldsymbol{s}))^{-\left(\frac{\boldsymbol{p}+1}{2}+1\right)} * (\boldsymbol{\sigma}^2(\boldsymbol{s}))^{-\boldsymbol{n}/2}|\boldsymbol{W}(\boldsymbol{s})|^{1/2}\mathbf{exp} \\ &\qquad \left\{-\frac{1}{2\boldsymbol{\sigma}^2(\boldsymbol{s})}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}))^T\boldsymbol{W}(\boldsymbol{s})(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{s}))\right\} \end{aligned} \tag{23}$$

$$\begin{aligned} &\propto |\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}(\sigma^2(s))^{-\left(\frac{p+1}{2}+1\right)} * (\sigma^2(s))^{-n/2}|\boldsymbol{W}(s)|^{1/2} \\ &\quad \exp\left\{-\frac{1}{2\sigma^2(s)}\left(\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right.\right. \\ &\qquad \left.\left.+(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s))^T\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s))\right)\right\} \end{aligned} \tag{24}$$

$$\begin{aligned} &\propto |\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s))^T\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s))\right\} * \\ &\qquad (\sigma^2(s))^{-\left(\frac{n+p+1}{2}\right)-1}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y})\right\} \end{aligned} \tag{25}$$

$$\begin{aligned} &\propto \left((\sigma^2)^{-\left(\frac{p}{2}\right)}|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}\exp\left\{-\frac{1}{2\sigma^2(s)}(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s))^T\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s))\right\}\right) * \\ &\qquad \left((\sigma^2(s))^{-\left(\frac{n+1}{2}\right)-1}\exp\left\{-\frac{1}{2\sigma^2(s)}(\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y})\right\}\right), \end{aligned} \tag{26}$$

where: $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X})^{-1}\boldsymbol{X}^T$.

Therefore, using Equation (26) as a starting point and assuming that the two distributions are independent of one another, the output of the posterior distribution can be represented as the multiplication of the two distributions [22], i.e., the marginal posterior distribution of $\boldsymbol{\beta}(s)$ is $MVN\left(\widehat{\boldsymbol{\beta}}(s), \sigma^2(s)(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X})^{-1}\right)$, while the marginal posterior distribution of $\sigma^2(s)$ is $IG\left(\frac{n+1}{2}, \frac{\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}}{2}\right)$.

4. The marginal posterior distribution of $\boldsymbol{\beta}(s)$

   Based on equation (26), we obtain that $\boldsymbol{\beta}(s)$ is $MVN\left(\widehat{\boldsymbol{\beta}}(s), \sigma^2(s)(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X})^{-1}\right)$:

   $$f(\boldsymbol{\beta}(s)|\boldsymbol{Y}) \propto (\sigma^2)^{-\left(\frac{p}{2}\right)}|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}\exp\left\{-\frac{1}{2\sigma^2(s)}\left(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s)\right)^T\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}\left(\boldsymbol{\beta}(s)-\widehat{\boldsymbol{\beta}}(s)\right)\right\}.$$

5. The marginal posterior distribution of $\sigma^2(s)$

   Based on Equation (26), the marginal posterior distribution of $\sigma^2(s)$ is $IG\left(\frac{n+1}{2}, \frac{\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}}{2}\right)$:

   $$f(\sigma^2(s)|\boldsymbol{Y}) \propto \left(\sigma^2(s)\right)^{-\left(\frac{n+1}{2}\right)-1}\exp\left\{-\frac{1}{2\sigma^2(s)}(\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y})\right\}.$$

4. CONCLUSIONS

Based on the results, Jeffreys' uninformative prior has been obtained, which is $|\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X}|^{1/2}\left(\sigma^2(s)\right)^{-(p+3)/2}$. Using the result, we have obtained a marginal posterior distribution for the $\boldsymbol{\beta}(s)$, which follows a normal multivariate distribution, that is, $\boldsymbol{\beta}(s)\sim MVN\left(\widehat{\boldsymbol{\beta}}(s), \sigma^2(s)(\boldsymbol{X}^T\boldsymbol{W}(s)\boldsymbol{X})^{-1}\right)$, whereas the marginal posterior distribution for σ^2 follows an inverse gamma distribution, that is, $\sigma^2(s)\sim IG\left(\frac{n+1}{2}, \frac{\boldsymbol{Y}^T\boldsymbol{W}(s)\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}}{2}\right)$.

As further research, it is necessary to follow up on some of the limitations of the results of this research, namely numerical simulations for several parameter conditions in the marginal posterior distributions of $\boldsymbol{\beta}(s)$ and $\sigma^2(s)$. Furthermore, to help readers better understand our work, it is necessary to provide an application in real cases, such as the Human Development Index (HDI) problem, which has several rigid assumptions in solving the model.

ACKNOWLEDGMENT

The authors appreciate the referees for their helpful comments on how to enhance this work. Partial funding for this research was also supported by FMIPA and Bengkulu University.

REFERENCES

[1] S. Astutik, A. B. Astuti, A. Efendi, Darmanto, D. Irsandy, and F. Y. D. A. S. Saniyawati, *Analisis Bayesian: Teori dan Aplikasi dengan R*. Malang: UB Press, 2023.

[2] S. T. Rachev, J. S. J. Hsu, B. S. Bagasheva, and F. J. Fabozzi, *Bayesian Methods in Finance*, 1st editio. John Wiley & Sons, Inc., Hoboken, New Jersey., 2008.

[3] A. S Fotheringham; C. Brunsdon; M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex: John Wiley & Sons Ltd, 2002.

[4] A. Kashki, M. Karami, R. Zandi, and Z. Roki, "Evaluation of the effect of geographical parameters on the formation of the land surface temperature by applying OLS and GWR, A case study Shiraz City, Iran," *Urban Clim.*, vol. 37, p. 100832, May 2021, doi: 10.1016/j.uclim.2021.100832.

[5] A. Nazarpour, G. Rostami Paydar, F. Mehregan, S. J. Hejazi, and M. A. Jafari, "Application of geographically weighted regression (GWR) and singularity analysis to identify stream sediment geochemical anomalies, case study, Takab Area, NW Iran," *J. Geochemical Explor.*, vol. 235, p. 106953, Apr. 2022, doi: 10.1016/j.gexplo.2022.106953.

[6] A. Ristea, O. Kounadi, and M. Leitner, "Geosocial media data as predictors in a GWR application to forecast crime hotspots," *Leibniz Int. Proc. Informatics, LIPIcs*, vol. 114, no. 56, pp. 1–7, 2018, doi: 10.4230/LIPIcs.GIScience.2018.56.

[7] Y.-C. Chiou, R.-C. Jou, and C.-H. Yang, "Factors affecting public transportation usage rate: Geographically weighted regression," *Transp. Res. Part A Policy Pract.*, vol. 78, pp. 161–177, Aug. 2015, doi: 10.1016/j.tra.2015.05.016.

[8] A. S. Fotheringham, R. Crespo, and J. Yao, "Exploring, modelling and predicting spatiotemporal variations in house prices," *Ann. Reg. Sci.*, vol. 54, no. 2, pp. 417–436, 2015, doi: 10.1007/s00168-015-0660-6.

[9] N. Subedi, L. Zhang, and Z. Zhen, "Bayesian geographically weighted regression and its application for local modeling of relationships between tree variables," *iForest - Biogeosciences For.*, vol. 11, no. 5, pp. 542–552, Oct. 2018, doi: 10.3832/ifor2574-011.

[10] J. P. LeSage, "A Family of Geographically Weighted Regression Models," 2004, pp. 241–264. doi: 10.1007/978-3-662-05617-2_11.

[11] I. G. N. M. Jaya and N. Sunengsih, "Bayesian Geographically Weighted Regression Dalam Pemodelan Angka Incidence Rate," *Euclid*, vol. 5, no. 1, p. 33, 2018, doi: 10.33603/e.v5i1.707.

[12] I. Sodikin, H. Pramoedyo, and S. Astutik, "Geographically weighted regression and bayesian geographically weighted regression modelling with adaptive gaussian kernel weight function on the poverty level in west Java Province," *Int. J. Humanit. Relig. Soc. Sci.*, vol. 2, no. 1, pp. 21–30, 2017.

[13] J. Geweke, "Bayesian Treatment of The Independent Student-t Linear Model," *J. Appl. Econom.*, vol. 8, no. September 1992, pp. S19–S40, 1993.

[14] Z. Ma, Y. Xue, and G. Hu, "Geographically Weighted Regression Analysis for Spatial Economics Data: A Bayesian Recourse," *Int. Reg. Sci. Rev.*, vol. 44, no. 5, pp. 582–604, Sep. 2021, doi: 10.1177/0160017620959823.

[15] R. van de Schoot *et al.*, "Bayesian statistics and modelling," *Nat. Rev. Methods Prim.*, vol. 1, no. 1, p. 1, Jan. 2021, doi: 10.1038/s43586-020-00001-2.

[16] F. A. Moala and G. Moraes, "Objective Prior Distributions to Estimate the Parameters of the Poisson-Exponential Distribution," *Rev. Colomb. Estadística*, vol. 46, no. 1, pp. 93–110, Jan. 2023, doi: 10.15446/rce.v46n1.95989.

[17] Y.-Y. Zhang, T.-Z. Rong, and M.-M. Li, "The Bayes Estimators of the Variance and Scale Parameters of the Normal Model With a Known Mean for the Conjugate and Noninformative Priors Under Stein's Loss," *Front. Big Data*, vol. 4, Jan. 2022, doi: 10.3389/fdata.2021.763925.

[18] M. Teimouri, "Fast Bayesian Inference for Birnbaum-Saunders Distribution," *Comput. Stat.*, vol. 38, no. 2, pp. 569–601, Jun. 2023, doi: 10.1007/s00180-022-01234-3.

[19] A. Ly, J. Verhagen, and E.-J. Wagenmakers, "Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology," *J. Math. Psychol.*, vol. 72, pp. 19–32, Jun. 2016, doi: 10.1016/j.jmp.2015.06.004.

[20] Gelman.A;Carlin;J.B;Stern.H.S;Dunson.D;Vehtari.A; Rubin.D, *Bayesian Data Analaysis*, 3rd Editio. 2014.

[21] B. Puza, *Bayesian methods for statistical analysis*. Acton ACT 2601, Australia: ANU eView, 2015.

[22] D. R. S. Saputro, F. Amalia, P. Widyaningsih, and R. C. Affan, "Parameter estimation of multivariate multiple regression model using bayesian with non-informative Jeffreys' prior distribution," *J. Phys. Conf. Ser.*, vol. 1022, no. 1, 2018, doi: 10.1088/1742-6596/1022/1/012002.