Geographically and temporally weighted multivariate generalised gamma regression for modelling three educational indicators in Central Java, Indonesia

Hasbi Yasin
Department of Statistics, Institut Teknologi Sepuluh Nopember,
Department of Statistics, Universitas Diponegoro, Indonesia
Email: 7003211002@student.its.ac.id
hasbiyasin@live.undip.ac.id

Purhadi
(corresponding author)
Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia
Email: purhadi@its.ac.id

Achmad Choiruddin
Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia,
Email: choiruddin@its.ac.id

Keywords:
skewed-positive distribution, spatio-temporal heterogeneity, GTWMGGR, educational indicators, multivariate generalized gamma

Goal 4 of the United Nations Sustainable Development Goals, which focuses on quality education, is relevant in the context of Central Java. With varying achievements concerning educational indicators across districts/cities and skewed-positive distribution, Central Java must address educational challenges and improve the quality of students' learning experiences. While many regression-based models have been introduced, no model has captured spatio-temporal variation for skewed-positive distributions. This study proposes a geographically and temporally weighted multivariate generalised gamma regression (GTWMGGR) model and applies it to model three educational indicators in Central Java, encompassing the mean years of schooling (MYS), school enrolment rate (SER) and gross enrolment rate (GER). We develop parameter estimation using the maximum likelihood method combined with Berndt–Hall–Hall–Hausman algorithm. In addition, we derive the hypothesis test procedure using the maximum likelihood ratio method. The results indicate that the GTWMGGR model parameters are significant at any given point., revealing eleven groups for MYS and GER indicators and nine groups for SER indicators. By identifying the specific educational contexts for each group, our study highlights the importance of developing tailored strategies and interventions to improve educational outcomes in Central Java.

Introduction

The geographically weighted regression (GWR) model is an extension of the linear regression model that can be applied to address spatial heterogeneity (Fotheringham et al. 2015, Fábián 2014, Bertus 2017, Safaralizadeh et al. 2024). In addition, the geographically and temporally weighted regression (GTWR) model is a valuable method for investigating spatio-temporal heterogeneity in a regression equation (Huang et al. 2010). Furthermore, Fotheringham et al. (2015) proposed an enhanced version of GTWR by introducing independent determination of the optimal spatial and temporal bandwidths, enabling a more comprehensive analysis of the temporal evolution of spatial patterns. While GTWR has been applied for research in various fields (Liu et al. 2017, Du et al. 2018, Zhang et al. 2019, Wang et al. 2020, Sifriyani et al. 2022), the model is restricted to (1) univariate cases and (2) a response variable with normal distribution. To extend GTWR, GWR models have been developed for non-normal univariate responses (da Silva−Rodrigues 2014, da Silva−de Oliviera Lima 2017, Putri et al. 2017, Yasin et al. 2022a, 2022b). Similarly, GWR models have been constructed to evaluate multivariate responses (Harini et al. 2012, Rahayu et al. 2020, Wenur et al. 2020, Chen et al. 2022, Palmí-Perales et al. 2023). However, to the best of our knowledge, no study has developed a local multivariate spatio-temporal model for skewed-positive distributions.

Generalised gamma (GG) distribution is a versatile continuous distribution that provides a more flexible framework than other distributions such as gamma, exponential, Erlang, Weibull, chi-squared, Rayleigh, log-normal and half-normal. It also has multiple benefits compared with normal and classical gamma distributions. With three parameters, its flexibility allows for a wider range of shapes and accommodates various skewness and kurtosis levels. GG can model heavy-tailed data and accommodate skewness better than normal distribution. The robustness of GG makes it suitable for skewed datasets. It can model various data types, including continuous, positive-valued and right-skewed data, which makes it a preferred choice in statistical applications (Stacy 1962, Sanchez–MacKenzie 2016, Shanker–Shukla 2016, Diantini–Purhadi-Choiruddin 2023). This study proposes the geographically and temporally weighted multivariate generalised gamma regression (GTWMGGR) model to explore the spatio-temporal modelling for multivariate generalised gamma (MGG) distributed responses. We thoroughly examine the parameter estimation process using maximum likelihood estimation (MLE) optimised by the Berndt–Hall–Hall–Hausman (BHHH) algorithm, and the model inference process involves the creation of a hypothesis test using the maximum likelihood ratio test (MLRT) method. Furthermore, we evaluate our proposed model in simulation study and in application to the educational indicators in Central Java, consisting of the mean years of schooling (MYS), gross enrolment rate (GER) and school enrolment rate (SER) indicators.

The remainder of this article is structured as follows. First, the MGGR model is described. This is followed by a detailed explanation of the GTWMGGR model, encompassing the methodology for estimating parameters in addition to hypothesis testing. Then, model validation in a simulation study is discussed. Next, the model's implementation on the educational indicator data for Central Java, Indonesia is examined. Finally, the relevant findings and proposes suggestions for future research endeavours are summarised.

Multivariate generalised gamma regression

Suppose $U_1, U_2, \cdots, U_K$ includes K random variables, each of which has GG distribution with λ and τ as the first and the second shape parameter, with θ_k as the k-th scale parameter and δ_k as the k-th location parameter, which can be written as $U_k \sim \mathrm{GG}(\lambda, \tau, \theta_k, \delta_k)$, $k = 1, 2, ..., K$. Therefore, if we define $Y_k = U_1 + U_2 + \cdots + U_k$, then the random vector of $Y_1, Y_2, \cdots, Y_K$ will be MGG distributed with the following probability density function (pdf) (Yasin et al. 2023):

$$f(\mathbf{y}|\boldsymbol{\Theta}) = \left(\frac{\tau}{\Gamma(\lambda)}\right)^K \frac{\exp\left(-\left(\frac{y_1-\delta_1}{\theta_1}\right)^{\tau} - \sum_{k=2}^{K}\left(\frac{y_k - y_{k-1} - \delta_k}{\theta_k}\right)^{\tau}\right)}{\left(\prod_{k=1}^{K}\theta_k\right)\left(\frac{y_1-\delta_1}{\theta_1}\right)^{1-\tau\lambda}\prod_{k=2}^{K}\left(\frac{y_k - y_{k-1} - \delta_k}{\theta_k}\right)^{1-\tau\lambda}}, \tag{1}$$

where $\boldsymbol{\Theta} = \left[\lambda\ \tau\ \boldsymbol{\theta}^T\ \boldsymbol{\delta}^T\right]^T$, $\mathbf{y} = \left[y_1\ y_1\ \cdots\ y_K\right]^T$, $\lambda, \tau > 0$, $\boldsymbol{\delta}^T = \left[\delta_1\ \delta_1 \cdots \delta_k \cdots \delta_K\right]$, $\delta_k \in \mathbb{R}$, $\boldsymbol{\theta}^T = \left[\theta_1\ \theta_2 \cdots \theta_k \cdots \theta_K\right]$, $\theta_k > 0$, $y_k > 0$, $k = 1, 2, \cdots, K$, $y_1 + \delta_k < y_k$; $k > 1$, $\delta_1 < y_1$ and $f(\mathbf{y}|\boldsymbol{\Theta}) = 0$ otherwise. The theoretical mean, variance, covariance and correlation are expressed in equations (2)–(5) (Yasin et al. 2023) as follows:

$$\mathrm{E}(Y_k) = \frac{\Gamma\left(\lambda + \frac{1}{\tau}\right)}{\Gamma(\lambda)}\sum_{r=1}^{k}\theta_r + \sum_{r=1}^{k}\delta_r, \tag{2}$$

$$\mathrm{Var}(Y_k) = \left(\frac{\Gamma\left(\lambda + \frac{2}{\tau}\right)}{\Gamma(\lambda)} - \left(\frac{\Gamma\left(\lambda + \frac{1}{\tau}\right)}{\Gamma(\lambda)}\right)^2\right)\sum_{r=1}^{k}\theta_r^2, \tag{3}$$

$$\mathrm{Cov}\left(Y_{k_1}, Y_{k_2}\right) = \mathrm{Var}\left(Y_{k_1}\right) \text{ for } k_1 < k_2, \text{ and} \tag{4}$$

$$\rho_{k_1,k_2} = \mathrm{Corr}\left(Y_{k_1}, Y_{k_2}\right) = \sqrt{\sum_{k=1}^{k_1}\theta_k^2 \Big/ \sum_{k=1}^{k_2}\theta_k^2}, \tag{5}$$

The scale parameter is used as a starting point for the MGGR model, and the log function is used as a connecting function. This implies that variations in the predictors directly correlate with variations in the scale parameter, whereas the other parameters are consistent for all data points (Yasin et al. 2024a). In the spatio-temporal data paradigm, the following equations express the matrix representing the response and predictor variables:

$$\mathbf{Y}_{(nL\times K)}=\left[\mathbf{y}_{11}^T \quad \mathbf{y}_{21}^T \quad \cdots \quad \mathbf{y}_{il}^T \quad \cdots \quad \mathbf{y}_{nL}^T\right], \tag{6}$$

$$\mathbf{X}_{(nL\times(p+1))}=\left[\mathbf{x}_{11}^T \quad \mathbf{x}_{21}^T \quad \cdots \quad \mathbf{x}_{il}^T \quad \cdots \quad \mathbf{x}_{nL}^T\right]^T, \tag{7}$$

where $\mathbf{y}_{il}^T=\left[y_{1il} \quad y_{2il} \quad \cdots \quad y_{kil} \quad \cdots \quad y_{Kil}\right]_{(1\times K)}$, $i=1,2,\cdots,n$, $l=1,2,\cdots,L$ and $\mathbf{x}_{il}^T=\left[1 \quad x_{1il} \quad x_{2il} \quad \cdots \quad x_{jil} \quad \cdots \quad x_{pil}\right]_{(1\times(p+1))}$. $\mathbf{y}_{il}$ and $\mathbf{x}_{il}$ are the vectors of response and predictors, respectively, at the *i*-th site in the *l*-th period. Therefore, using equation (2), the mean of Y_{kil} can be written as follows:

$$\mathrm{E}\left(Y_{kil}\right)=\frac{\Gamma\left(\lambda+\frac{1}{\tau}\right)}{\Gamma(\lambda)}\sum_{r=1}^{k}\theta_r\left(\mathbf{x}_{il}\right)+\sum_{r=1}^{k}\delta_r, \tag{8}$$

where $\theta_1\left(\mathbf{x}_{il}\right)=\dfrac{\Gamma(\lambda)}{\Gamma\left(\lambda+\frac{1}{\tau}\right)}\left(\exp\left(\mathbf{x}_{il}^T\boldsymbol{\beta}_1\right)-\delta_1\right)$, and (9)

$$\theta_k\left(\mathbf{x}_{il}\right)=\frac{\Gamma(\lambda)}{\Gamma\left(\lambda+\frac{1}{\tau}\right)}\left(\exp\left(\mathbf{x}_{il}^T\boldsymbol{\beta}_k\right)-\exp\left(\mathbf{x}_{il}^T\boldsymbol{\beta}_{k-1}\right)-\delta_{ki}\right), k=2,3,\cdots,K. \tag{10}$$

Therefore, considering equations (8)–(10), the MGGR model can be defined by modelling the mean of the random variable as follows:

$$\mathrm{E}\left(Y_{kil}\right)=\exp\left(\mathbf{x}_{il}^T\boldsymbol{\beta}_k\right), \tag{11}$$

where $\boldsymbol{\beta}_k=\left[\beta_{0k}\ \beta_{1k}\cdots\beta_{jk}\cdots\beta_{pk}\right]^T$ reflects the regression coefficients obtained from *p* predictors. Therefore, the vector of the MGGR parameter can be written as $\boldsymbol{\Theta}_R=\left[\boldsymbol{\beta}_1^T\ \boldsymbol{\beta}_2^T\cdots\boldsymbol{\beta}_k^T\cdots\boldsymbol{\beta}_K^T\ \lambda\ \tau\ \boldsymbol{\delta}^T\right]^T$. The model's parameter estimation can be obtained employing the MLE approach, which is optimised using the BHHH algorithm. Hypothesis testing of the regression parameters simultaneously uses the MLRT and partially uses the Wald test (Yasin et al. 2024a).

Geographically and temporally weighted multivariate generalised gamma regression (GTWMGGR)

GTWMGGR can be considered an extended version of MGGR for considering spatio-temporal heterogeneity similar to GTWR for linear regression. To account for spatio-temporal heterogeneity, we mimic the idea of GTWR by defining the spatio-temporal weighting matrix with components that have a spatio-temporal kernel and incorporating spatio-temporal distance with a certain bandwidth. The spatio-temporal weighting function of the i^*-th site against the i-th site at the l-th period can be determined as follows:

$$w_{ii^*}^{l} = \phi_S\left(d_{ii^*}^{S}, h_{Sl}\right) \times \phi_T\left(d_{ii^*}^{l}, h_T\right), \tag{12}$$

where ϕ_S and ϕ_T are spatial and temporal kernel functions, respectively, $d_{ii^*}^{S}$ represents the spatial distance across i and i^* (determined using the Euclidean distance formula), $d_{ii^*}^{l}$ is the temporal distance measured using the time lag of the l-th period to the current time and h_{Sl} and h_T are spatial and temporal bandwidths, respectively. Thus, the spatio-temporal weight matrix for the i^*-th location in L data periods includes L components, where each component represents the l-th period as follows:

$$\mathbf{W}_{i^*}^{(ST)} = diag\left(\mathbf{W}_{i^*1}^{(ST)}, \mathbf{W}_{i^*2}^{(ST)}, \cdots, \mathbf{W}_{i^*l}^{(ST)}, \cdots, \mathbf{W}_{i^*L}^{(ST)}\right), \tag{13}$$

where the weighting matrix for the i^*-th location in the l-th period is as follows:

$$\mathbf{W}_{i^*l}^{(ST)} = diag\left(w_{1i^*}^{l}, w_{2i^*}^{l}, \cdots, w_{ii^*}^{l}, \cdots, w_{ni^*}^{l}\right). \tag{14}$$

The weight assigned to each observation is decided by considering its spatial and temporal proximity to others. This dual consideration of space and time allows for a nuanced understanding of how different locations and moments in time contribute to the overall influence on a given response variable. In essence, the weight value can be expressed because of the combined spatial and temporal relationships captured by the chosen kernel function, providing a framework for simultaneously analysing the dynamics of a phenomenon over space and time. The weight matrix components are flexible with the kernel function used. Using a Gaussian kernel function, the spatio-temporal weights are computed as follows:

$$w_{ii^*}^{l} = \exp\left(-\frac{1}{2}\left(\frac{d_{ii^*}^{S}}{h_{Sl}}\right)^2\right) \times \exp\left(-\frac{1}{2}\left(\frac{d_{ii^*}^{l}}{h_T}\right)^2\right). \tag{15}$$

Applying the spatio-temporal weighting matrix to the MGGR model will yield distinct parameters for each site; therefore, the general GTWMGGR model framework is expressed as follows:

$$\mu_{kil} = \mathrm{E}\left(Y_{kil}\right) = \exp\left(\mathbf{x}_{il}^{\boldsymbol{T}} \boldsymbol{\beta}_{ki}\right), \tag{16}$$

where $\boldsymbol{\beta}_{ki}^{T}=\left[\beta_{0ki}\ \beta_{1ki}\cdots\beta_{jki}\cdots\beta_{pki}\right]_{((p+1)\times 1)}$ represents the regression coefficients, assuming a spatio-temporal heterogeneity of impact. Since parameters are estimated at each point, we can generate the pdf of the GTWMGGR model by modifying the pdf of the MGGR model through adding a location index to each model parameter. Therefore, the pdf of the GTWMGGR model with the MGG-distributed response variable ($y_{1il}, y_{2il}, \cdots, y_{Kil}$) at the i-th site can be obtained as follows:

$$f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i}\right.\right)=\left(\prod_{k=1}^{K}\frac{\tau_i}{\theta_{ki}\left(\mathbf{x}_{il}\right)\Gamma\left(\lambda_i\right)}\right)\frac{\exp\left(-\left(\frac{y_{1il}-\delta_{1i}}{\theta_{1i}\left(\mathbf{x}_{il}\right)}\right)^{\tau_i}-\sum_{k=2}^{K}\left(\frac{y_{kil}-y_{1il}-\delta_{ki}}{\theta_{ki}\left(\mathbf{x}_{il}\right)}\right)^{\tau_i}\right)}{\left(\frac{y_{1il}-\delta_{1i}}{\theta_{1i}\left(\mathbf{x}_{il}\right)}\right)^{1-\tau_i\lambda_i}\prod_{k=2}^{K}\left(\frac{y_{kil}-y_{1il}-\delta_{ki}}{\theta_{ki}\left(\mathbf{x}_{il}\right)}\right)^{1-\tau_i\lambda_i}}, \quad (17)$$

where $$\theta_{1i}\left(\mathbf{x}_{il}\right)=\frac{\Gamma\left(\lambda_i\right)}{\Gamma\left(\lambda_i+\frac{1}{\tau_i}\right)}\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i}\right)-\delta_{1i}\right), \quad (18)$$

$$\theta_{ki}\left(\mathbf{x}_{il}\right)=\frac{\Gamma\left(\lambda_i\right)}{\Gamma\left(\lambda_i+\frac{1}{\tau_i}\right)}\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i}\right)-\delta_{ki}\right), k=2,3,\cdots,K, \quad (19)$$

$\boldsymbol{\Theta}_{GTR}^{i}=\left[\boldsymbol{\beta}_{1i}^{T}\ \boldsymbol{\beta}_{2i}^{T}\cdots\boldsymbol{\beta}_{ki}^{T}\cdots\boldsymbol{\beta}_{Ki}^{T}\ \lambda_i\ \tau_i\ \boldsymbol{\delta}_{i}^{T}\right]^{T}$ represents the GTWMGGR model parameters at the i-th point, $\lambda_i>0$ $\tau_i>0$, $\boldsymbol{\delta}_{i}^{T}=\left[\delta_{1i}\ \ \delta_{2i}\ \cdots\ \delta_{ki}\ \cdots\ \delta_{Ki}\right]_{(1\times K)}$, $\delta_{ki}\in\mathbb{R}$, $y_{1il}+\delta_{ki}<y_{kil}; k>1$, $\delta_{1i}<y_{1il}$ and $f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i}\right.\right)=0$ for others.

GTWMGGR model parameter estimation

The estimator of GTWMGGR parameters is partially obtained at every location by involving spatio-temporal weighting elements. Suppose that $\boldsymbol{\Theta}_{GTR}^{i*}$ is the parameter of the GTWMGGR model at the i*-th location to be estimated. The log-likelihood function is constructed by multiplying the spatio-temporal weights w_{ii*}^{l} as follows:

$$L\left(\boldsymbol{\Theta}_{GTR}^{i*}\right)=\sum_{l=1}^{L}\sum_{i=1}^{n}L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i*}\right)=\sum_{l=1}^{L}\sum_{i=1}^{n}w_{ii*}^{l}\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right). \quad (20)$$

The estimator of $\boldsymbol{\Theta}_{GTR}^{i*}$ can be derived using the MLE approach by setting the first derivative of the likelihood function ($L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i*}\right)$) concerning $\boldsymbol{\Theta}_{GTR}^{i*}$ equal to zero and ensuring that the second derivative of $L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i*}\right)$ concerning $\boldsymbol{\Theta}_{GTR}^{i*}$ is a negative definite matrix. Due to the differentiation of $L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i*}\right)$ with respect to all elements of $\boldsymbol{\Theta}_{GTR}^{i^*}$ not yielding a closed form, the uniqueness of our technique is the use of the BHHH algorithm to tackle the defined challenge efficiently. A numerical procedure that combines a likelihood criterion to fit the model optimises the criterion in this iteration method. The method involves iteratively updating the parameters by using the inverse of the expected Fisher information matrix, which is defined by the gradient function (Berndt et al. 1974, Greene 2008, Wooldridge 2010). This method was adopted in practice because it eliminates the need to calculate the second derivative to generate the Hessian matrix. The approximate Hessian matrix can be computed by summing the gradient vectors of each observation (Rahayu et al. 2020, Wenur et al. 2020). Equations (21) and (22) express the GTWMGGR parameters' gradient and approximate Hessian matrix.

$$\mathbf{g}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)=\sum_{l=1}^{L}\sum_{i=1}^{n}\mathbf{g}_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right), \tag{21}$$

$$\mathbf{H}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)=-\sum_{l=1}^{L}\sum_{i=1}^{n}\mathbf{g}_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)\mathbf{g}_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)^{T}, \tag{22}$$

where $\mathbf{g}_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)$ denotes the gradient of $\boldsymbol{\Theta}_{GTR}^{i^*}$ at the i-th site and l-th period that is presented in equation (23), see in Appendix A1 for more detailed expressions of the gradient and approximated Hessian matrix.

$$\mathbf{g}_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)=\left[\frac{\partial L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)}{\partial\boldsymbol{\beta}_{1i^*}^{T}}\cdots\frac{\partial L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)}{\partial\boldsymbol{\beta}_{Ki^*}^{T}}\frac{\partial L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)}{\partial\lambda_{i^*}}\frac{\partial L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)}{\partial\tau_{i^*}}\frac{\partial L_{il}\left(\boldsymbol{\Theta}_{GTR}^{i^*}\right)}{\partial\boldsymbol{\delta}_{i^*}^{T}}\right]^{T} \tag{23}$$

This estimation process is repeated at each location to estimate the overall GTWMGGR model parameters or $\hat{\boldsymbol{\Theta}}_{GTR}=\left[\hat{\boldsymbol{\Theta}}_{GTR}^{1}\,\hat{\boldsymbol{\Theta}}_{GTR}^{2}\cdots\hat{\boldsymbol{\Theta}}_{GTR}^{i^*}\cdots\hat{\boldsymbol{\Theta}}_{GTR}^{n}\right]^{T}$. Table 1 presents the steps for obtaining the $\hat{\boldsymbol{\Theta}}_{GTR}$.

Table 1

Algorithm for parameter estimation of the GTWMGGR model

Input: data $\mathbf{y}_{il}$, $\mathbf{x}_{il}$ where $i = 1,2,\cdots,n$ and $l = 1,2,\cdots,L$ as shown in equations (6) and (7)
geographical coordinates $\mathbf{u}_i = (u_i, v_i)$, time period l
spatial bandwidth h_{Sl}
temporal bandwidth h_T
for $i* = 1,2,\cdots,n$ **do**
Compute the spatial distance d_{ii*}^{S} where $i* = 1,2,\cdots,n$
for $l = 1,2,\cdots,L$ **do**
Compute the temporal distance $d_{ii*}^{l} = l - 1$ where $l = 1,2,\cdots,L$
Compute the spatio-temporal weight matrix $\mathbf{W}_{i*}^{(ST)}$
Initialise $\hat{\mathbf{\Theta}}_{GTR}^{i*(0)} = \hat{\mathbf{\Theta}}_R$, $\varepsilon = 10^{-8}$ and maximum iterations of $M = 150$
Define iteration $m = 0$
Compute the gradient $\mathbf{g}\left(\hat{\mathbf{\Theta}}_{GTR}^{i*(0)}\right)$ using equation (21)
Compute the Hessian matrix approximation $\mathbf{H}\left(\hat{\mathbf{\Theta}}_{GTR}^{i*(0)}\right)$ using equation (22)
while $\left\|\hat{\mathbf{\Theta}}_{GTR}^{i*(m+1)} - \hat{\mathbf{\Theta}}_{GTR}^{i*(m)}\right\| \geq \varepsilon$ **or** $m < M$ **do**
$\hat{\mathbf{\Theta}}_{GTR}^{i*(m+1)} = \hat{\mathbf{\Theta}}_{GTR}^{i*(m)} - \mathbf{H}^{-1}\left(\hat{\mathbf{\Theta}}_{GTR}^{i*(m)}\right)\mathbf{g}\left(\hat{\mathbf{\Theta}}_{GTR}^{i*(m)}\right)$
$m = m + 1$
end while
end for
end for

After we obtain the estimated GTWMGGR model parameters, the log-likelihood model can be calculated by summing up the log-likelihood of each observation based on respective parameters as follows:

$$L\left(\mathbf{\Theta}_{GTR}\right) = \sum_{i*=1}^{n} L\left(\mathbf{\Theta}_{GTR}^{i*}\right). \tag{24}$$

Testing the GTWMGGR model hypothesis

GTWMGGR parameters are tested simultaneously using hypothesis testing with the following hypothesis.

H0: $\beta_{jki} = 0$, for each $j = 1,2,\cdots,p$, $k = 1,2,\cdots,K$ and $i = 1,2,\ldots,n$,

H1: at least one of $\beta_{jki} \neq 0$.

Define Ω_{GTR} as the set of GTWMGGR model parameters under population and ω_{GTR} as the set of GTWMGGR model parameters under H_0, denoted by

$$\Omega_{GTR} = \{\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{21}, \cdots, \boldsymbol{\beta}_{ki}, \cdots, \boldsymbol{\beta}_{Kn}, \lambda_1, \cdots, \lambda_n, \tau_1, \cdots, \tau_n, \boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_n\} \text{ and}$$

$$\omega_{GTR} = \{\beta_{\omega 011}, \beta_{\omega 021}, \cdots, \beta_{\omega 0ki}, \cdots, \beta_{\omega 0Kn}, \lambda_{\omega 1}, \cdots, \lambda_{\omega n}, \tau_{\omega 1}, \cdots, \tau_{\omega n}, \boldsymbol{\delta}_{\omega 1}, \cdots, \boldsymbol{\delta}_{\omega n}\}.$$

In addition, $\hat{\Omega}_{GTR}$ and $\hat{\omega}_{GTR}$ are estimated using the MLE method as outlined in algorithm 1, which is employed using the MLRT method. The test statistics are stated as follows:

$$G^2_{GTR} = 2\left(L\left(\hat{\Omega}_{GTR}\right) - L\left(\hat{\omega}_{GTR}\right)\right), \tag{25}$$

where $L\left(\hat{\Omega}_{GTR}\right)$ and $L\left(\hat{\omega}_{GTR}\right)$ express the log-likelihood of the GTWMGGR using the respective estimators, under population and H_0, see equation (24). In Appendix A2, we prove that the test statistic G^2_{GTR} asymptotically follows the chi-squared distribution with $K\,\mathrm{tr}(\mathbf{S})$ degrees of freedom, where $\mathbf{S}$ expresses the GTWMGGR projection matrix in equation (26). Therefore, the null hypothesis (H_0) is rejected when $G^2_{GTR} > \chi^2_{(1-\alpha);K\,\mathrm{tr}(\mathbf{S})}$, where α is the significance level.

$$\mathbf{S}_{(nL\times nL)} = \begin{bmatrix} \mathbf{s}^T_{11} & \mathbf{s}^T_{21} & \cdots & \mathbf{s}^T_{il} & \cdots & \mathbf{s}^T_{nL} \end{bmatrix}^T, \tag{26}$$

where $\mathbf{s}^T_{il} = \left[\mathbf{x}^T_{il}\left(\mathbf{X}^T_l \mathbf{W}^{(ST)}_{il} \mathbf{X}_l\right)^{-1} \mathbf{X}^T_l \mathbf{W}^{(ST)}_{il}\right]_{(1\times nL)}$, and $\mathbf{X}_l$ is the predictor matrix for the lth period observation of size $(n\times(p+1))$. (The distribution of the statistical test in equation (25) is presented in Appendix A2.)

The hypothesis for testing the GTWMGGR model parameters partially is H_0: $\beta_{jki} = 0$ versus H_1: $\beta_{jki} \neq 0$, $j = 1, 2, \cdots, p$, $k = 1, 2, \cdots, K$ and $i = 1, 2, \ldots, n$. The hypothesis test statistics are obtained as follows:

$$Z_{jki} = \frac{\hat{\beta}_{jki}}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)}} \xrightarrow[n\to\infty]{d} N(0,1), \tag{27}$$

where $SE\left(\hat{\beta}_{jki}\right) = \sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)}$, and $\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)$ is the main diagonal element corresponding to $\hat{\beta}_{jki}$ of the matrix $-\left[\mathbf{H}\left(\hat{\boldsymbol{\Theta}}^i_{GTR}\right)\right]^{-1}$. Therefore, H_0 is rejected when $\left|Z_{jki}\right| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ represents the standard normal distribution quantile (the proof of equation (27) is presented in Appendix A3).

Finally, we conduct hypothesis testing to examine the significance of spatio-temporal effects by comparing the GTWMGGR model with the MGGR model using the following hypothesis:

H_0: $\beta_{jki} = \beta_{jk}; \forall j = 0,1,2,\cdots,p; k = 1,2,\cdots,K; i = 1,2,...,n$,

H_1: at least one of $\beta_{jki} \neq \beta_{jk}$.

The hypotheses' test statistics are obtained as follows:

$$F_{GTR} = \left(G_{\mathrm{R}}^2 \big/ pK \right) \Big/ \left(G_{GTR}^2 \big/ K\,\mathrm{tr}(\mathbf{S}) \right), \tag{28}$$

where G_{R}^2 is the test statistic used in the simultaneous hypothesis testing of the significance of the MGGR parameters. It is recognised that G_{R}^2 is chi-square distributed with $df = pK$ (Yasin et al. 2024a), and G_{GTR}^2 is also chi-square distributed with $K\,\mathrm{tr}(\mathbf{S})$. Therefore, it is straightforward to prove that the test statistic F_{GTR} is F-distributed with degrees of freedom pK and $K\,\mathrm{tr}(\mathbf{S})$; hence, H_0 is rejected if $F_{GTR} > F_{(1-\alpha);pK,K\mathrm{tr}(\mathbf{S})}$, where $F_{(1-\alpha);pK,K\mathrm{tr}(\mathbf{S})}$ represents the F distribution's $(1-\alpha)$-th quantile with pK and $K\,\mathrm{tr}(\mathbf{S})$ degrees of freedom. As noted by a reviewer, the effective number of parameters may be underestimated since the hat matrix S does not include the local shape and scale parameters. Further development based on the technique developed by da Silva–Rodrigues (2014) and da Silva–de Oliviera Lima (2017) is left for future research endeavours.

Remark: GTWMGGR is a spatio-temporal model developed using a spatio-temporal weight matrix. Our model is a more general case that can also cover the following spatial and temporal models.

1. The geographical model (GWMGGR) can be used when the weight matrix is only constructed based on the spatial heterogeneity effect (Yasin et al. 2024b). In this scenario, the weight matrix is solely influenced by spatial bandwidth and distance as follows:
$$\mathbf{W}_{i^*}^{(S)} = diag\left(\mathbf{W}_{i^*1}^{(S)}, \mathbf{W}_{i^*2}^{(S)}, \cdots, \mathbf{W}_{i^*l}^{(S)}, \cdots, \mathbf{W}_{i^*L}^{(S)} \right), \tag{29}$$
where $\mathbf{W}_{i^*l}^{(S)} = diag\left(w_{1i^*}^l, w_{2i^*}^l, \cdots, w_{ii^*}^l, \cdots, w_{ni^*}^l \right)$ and $w_{ii^*}^l = \exp\left(-\frac{1}{2} \left(\frac{d_{ii^*}^S}{h_{Sl}} \right)^2 \right)$.
2. The temporal model (TWMGGR) can be used when the weight matrix is only constructed based on the temporal heterogeneity effect. In this scenario, the weight matrix is only affected by temporal bandwidth and distance as follows:
$$\mathbf{W}^{(T)} = diag\left(\mathbf{W}_1^{(T)}, \mathbf{W}_2^{(T)}, \cdots, \mathbf{W}_l^{(T)}, \cdots, \mathbf{W}_L^{(T)} \right), \tag{30}$$
where $\mathbf{W}_l^{(T)} = diag\left(w_{1i^*}^l, w_{2i^*}^l, \cdots, w_{ii^*}^l, \cdots, w_{ni^*}^l \right)$ and $w_{ii^*}^l = \exp\left(-\frac{1}{2} \left(\frac{d_{ii^*}^l}{h_T} \right)^2 \right)$.

Simulation study

Simulation design

To validate the effectiveness of the proposed model, we apply the GTWMGGR model on several simulated datasets. Due to limitations in the generating spatio-temporal datasets, we only test a spatial simulation using the GWMGGR model. The GWMGGR model is a GTWMGGR model that uses the weighting matrix in equation (29). The datasets have known characteristics and exhibit different degrees of spatial heterogeneity. The spatial arrangement is represented by a 15 x 15 lattice with a regular pattern. We established three different parameter surfaces for a single covariate, each of which respectively exhibits zero, medium and high levels of spatial heterogeneity to investigate the effect of these different levels on the model's effectiveness. These surfaces were designed according to the following rules (Fotheringham et al. 2017).

Simulation design 1 (zero spatial heterogeneity):

$\beta_{01i} = 2.5$, $\beta_{02i} = 3.5$ and $\beta_{03i} = 4.0$,

$\beta_{11i} = 0.0044$, $\beta_{12i} = 0.0011$ and $\beta_{13i} = 0.0013$.

Simulation design 2 (low spatial heterogeneity):

$\beta_{01i} = 2.5$, $\beta_{02i} = 3.5$ and $\beta_{03i} = 4.0$,

$$\beta_{11i} = \log\left(1+\left(\frac{1}{12}\right)(u_i+v_i)\right)\Big/100,\ \beta_{12i} = \log\left(1+\left(\frac{1}{12}\right)(u_i+v_i)\right)\Big/75 \text{ and}$$

$$\beta_{13i} = \log\left(1+\left(\frac{1}{12}\right)(u_i+v_i)\right)\Big/85.$$

Simulation design 3 (high spatial heterogeneity):

$\beta_{01i} = 2.5$, $\beta_{02i} = 3.5$ and $\beta_{03i} = 4.0$,

$$\beta_{11i} = \log\left(1+\left(\frac{1}{324}\right)\left(36-\left(6-\frac{u_i}{2}\right)^2\right)\left(36-\left(6-\frac{v_i}{2}\right)^2\right)\right)\Big/100,$$

$$\beta_{12i} = \log\left(1+\left(\frac{1}{324}\right)\left(36-\left(6-\frac{u_i}{2}\right)^2\right)\left(36-\left(6-\frac{v_i}{2}\right)^2\right)\right)\Big/75 \text{ and}$$

$$\beta_{13i} = \log\left(1+\left(\frac{1}{324}\right)\left(36-\left(6-\frac{u_i}{2}\right)^2\right)\left(36-\left(6-\frac{v_i}{2}\right)^2\right)\right)\Big/85,$$

where v_i and u_i represent the vertical and horizontal coordinates, respectively, which increase in increments of 1. The algorithm for model simulation is presented in Table 2.

Table 2

GWMGGR simulation algorithm

Define: n = 15x15, the spatial layout is structured as a consistent 15 x 15 grid using geographical coordinates (u_i, v_i).

$K = 3$ responses, and $p = 1$ predictor.

Assume: $\boldsymbol{\beta}_{1i}, \boldsymbol{\beta}_{2i}$ and $\boldsymbol{\beta}_{3i}$ (the true regression coefficients for three responses) following the three simulation designs.

$\lambda = 1.5, \tau = 3.75$ and $\boldsymbol{\delta}^T = [3, 10, 8]$ as the MGGD parameters

$X_1 \sim U(5, 30)$ as predictor

Data generation

for $i = 1, 2, \cdots, n$ **do**

generate $\mathbf{x}_i = \begin{bmatrix} 1 & x_{1i} \end{bmatrix}^T$

calculate $\theta_{1i}(\mathbf{x}_i) = \dfrac{\Gamma(\lambda)}{\Gamma\left(\lambda + \dfrac{1}{\tau}\right)} \left(\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_{1i}\right) - \delta_1\right)$,

$$\theta_{ki}(\mathbf{x}_i) = \frac{\Gamma(\lambda)}{\Gamma\left(\lambda + \frac{1}{\tau}\right)} \left(\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_{ki}\right) - \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_{(k-1)i}\right) - \delta_k\right), k = 2, 3$$

generate $\mathbf{y}_i \sim MGG(\lambda, \tau, \boldsymbol{\theta}_i, \boldsymbol{\delta})$, where $\boldsymbol{\theta}_i = \begin{bmatrix} \theta_{1i}(\mathbf{x}_i) & \theta_{2i}(\mathbf{x}_i) & \theta_{3i}(\mathbf{x}_i) \end{bmatrix}^T$

end for

Apply GWMGGR to $\mathbf{y}$ on x_1

Simulation results

Table 3 presents the average of the important statistics for MGGR against GWMGGR, considering three spatial heterogeneity levels. The results reveal that the optimal bandwidth of the spatial model (GWMGGR) will decrease when dealing with data that exhibits significant spatial heterogeneity. Consequently, the effective parameters will rise substantially, causing the model parameters to exhibit more significant variability. Then, the MGGR model's log-likelihood decreases as the amount of spatial heterogeneity rises, indicating less accuracy. In contrast, the precision accuracy of the GWMGGR model will be enhanced. This is demonstrated by the increasing gap in deviation between the two models. Analysing the similarity test of the MGGR against GWMGGR models reveals that as the degree of geographical heterogeneity rises, the F-test statistic value increases and the p-value lowers dramatically. Therefore, the GWMGGR model surpasses the MGGR model in effectively accommodating spatial heterogeneity. It is crucial to recognise that estimating the parameters of the GWMGGR model is quite time-consuming, contingent upon the level of geographical heterogeneity. For this reason, we only replicated the simulation process 30 times for each simulation design. Therefore, for future research, improving the likelihood function optimisation method using a faster computational method that is non-gradient based is recommended.

Table 3

Simulation result

Average of statistics		Spatial heterogeneity level		
		zero	low	high
MGGR	log-likelihood	−1,635.237	−1,837.094	−1,968.990
	deviance	3,270.474	3,674.187	3,937.979
GWMGGR	log-likelihood	−1,634.582	−1,776.496	−1,827.781
	deviance	3,269.266	3,552.992	3,655.562
Overall	difference of deviance	1.209	121.195	282.417
	optimum bandwidth	19.775	2.900	1.986
	n-eff	6.582	39.355	72.189
	F-test	2.088	11.100	15.723
	p-value of F-test	1.962×10^{-1}	1.868×10^{-4}	2.223×10^{-5}
	computation time (in hours)	0.190	0.779	1.248

Table 4

The accuracy of parameter estimation

Parameter	Spatial heterogeneity level					
	zero		low		high	
	MSE	averaged bias	MSE	averaged bias	MSE	averaged bias
β_{01}	0.003241	0.048593	0.004382	0.031957	0.009459	0.045262
β_{11}	0.000002	0.000031	0.000012	0.001369	0.000022	0.001737
β_{02}	0.000337	0.010374	0.001869	0.025163	0.004082	0.032791
β_{12}	0.000000	0.000023	0.000005	0.001202	0.000015	0.002511
β_{03}	0.002490	0.048080	0.002820	0.043499	0.003236	0.034615
β_{13}	0.000001	0.000014	0.000004	0.000950	0.000011	0.002196
Average	0.001012	0.017852	0.001515	0.017357	0.002804	0.019852

We evaluated the precision of the parameter estimates by quantifying the mean square error (MSE) and averaged bias of each estimated parameter, as shown in Table 4. The simulation results based on the MSE of each parameter demonstrate that the proposed algorithm effectively estimates parameters close to their actual values. Furthermore, the bias of each parameter also indicates that the average estimated parameter is not much different from the actual parameter. However, the average MSE tends to rise significantly as spatial heterogeneity grows because parameter variability will increase at higher levels of spatial heterogeneity. Nevertheless, the results show that our proposed method consistently yields precise parameter estimates for the GWMGGR framework with low bias.

Application in educational indicators

Next, we use published data from the BPS Central Java Province spanning 2017 to 2021, encompassing information from 35 districts/cities [1]. Table 5 shows that the average MYS is 7.77 years, with respective minimum and maximum of 6.18 years (Brebes district in 2017) and 10.90 years (Surakarta city in 2021). This indicates that the average education level of individuals aged 25 years and older is equivalent to completing only the first year of junior high school or dropping out in the second year, which is still far from the target of 9 years of compulsory education, which only 17% of districts and cities have achieved. On average, SER is 71.50, with minimum and maximum values of 49.56 (Brebes district in 2018) and 91.39 (Magelang city in 2019). This indicates that the average school enrolment rate for 16–18-year-olds in each district or city is approximately 71.50%, which falls short of the national target. Finally, the GER average is 85.73, with minimum and maximum values of 52.98 (Wonosobo district in 2017) and 121.91 (Salatiga city in 2019). This indicates that in each district or city, only approximately 85.73% of the population aged 16–18 is enrolled in senior high school on average, which is below the national goal of 88.39% for GER [3].

Table 5

Research data descriptive statistics

Research variable	Mean	Standard deviation	Minimum	Maximum
Mean years of schooling (Y_1) (years)	7.773	1.212	6.180	10.900
School enrolment rate of 16–18-year-olds (Y_2) (%)	71.501	9.011	49.560	91.390
Gross enrolment rate of high school level (Y_3) (%)	85.734	14.941	52.980	121.910
Per capita GRDP (X_1) (million rupiahs)	28.386	17.966	12.370	87.360
Proportion of poverty (X_2) (%)	11.245	3.673	3.980	20.320
Gender ratio (X_3)	99.373	2.306	93.900	103.940
Percentage of households with access to proper sanitation (X_4) (%)	77.954	17.292	9.240	98.070
Labour force participation rate (X_5) (%)	69.369	3.153	58.730	76.600
Student–teacher ratio (X_6)	18.063	2.934	13.000	39.000

Table 6

Correlation among response variables

Response	SER	GER
MYS	0.6428	0.4508
SER		0.5943

We calculate the correlation matrix between response variables using equation (5) and the results are presented in Table 6. The results reveal a high degree of correlation among response variables, as shown by testing the correlation using the Bartlett sphericity test approach (Bartlett 1951). The test statistic 169.933 (p-value = 0.000) is obtained based on the correlation matrix. Therefore, it is reasonable to conclude that all responses are significantly correlated, and it is feasible to conduct analyses using a multivariate regression approach.

Figure 1

Histogram of response variables

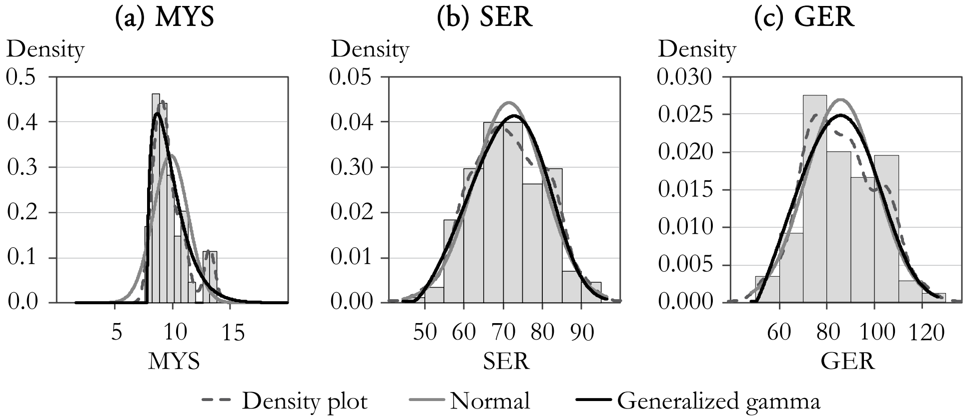


Table 7

Kolmogorov–Smirnov (KS) distribution test

Response	Distribution	Test statistic	p-value	Log-likelihood
MYS	normal	0.1552	0.0004*	−281.4215
	generalised gamma	0.0664	0.4241	−247.4655
SER	normal	0.0572	0.6164	−632.5494
	generalised gamma	0.0510	0.7523	−630.9296
GER	normal	0.0708	0.3439	−721.0267
	generalised gamma	0.0574	0.6121	−718.7643

Note: * indicates significance at $\alpha = 5\%$.

Furthermore, we test the distribution's goodness of fit using the Kolmogorov–Smirnov (KS) test (Conover 1971). Null hypothesis: the empirical data adheres to the GG distribution; alternative hypothesis: the observed data deviates from the GG distribution. The pdf of the GG distribution in this study refers to Yasin et al. (2023). Table 7 shows all response variables, univariately, under the GG distribution, revealing p-values that are all more significant than the 5% level. This is reinforced by the log-likelihood value, which is larger than the log-likelihood of the normal distribution, and the visualisation of the histogram, which is asymmetrical and skewed to the right (Figure 1). The test yields a statistical score of 0.052564 for the multivariate response, with a corresponding p-value of 0.7189. The results indicate that the null hypothesis should not be rejected, suggesting that the SER, GER and MYS conform to the MGG distribution (Yasin et al. 2023). These findings confirm that regression modelling based on the MGG distribution is highly recommended for this case because it more optimally accommodates information from the variation of the response variable compared with regression using normal distribution.

Modelling educational indicators using MGGR

Initially, we employ the MGGR model to determine the predictor variables that have a significant impact on the response variable. Table 8 presents the MGGR model parameters, which include three response variables and six predictor variables. At a 5% significance level, we can infer that all predictors have a substantial effect on Y_1, with the exception of variable X_5, the factors that significantly affect Y_2 are variables X_3, X_4 and X_6, and Y_3 is influenced substantially by variables X_3, X_4 and X_5. The model's relevance can be evaluated simultaneously by employing Wilk's likelihood ratio from the MLRT (Yasin et al. 2024a). The test score is 337.433 and the p-value is 0.0000, indicating that the six predictors significantly impact the response variables collectively.

Table 8

Coefficients of MGGR model

Parameter	Y_1		Y_2		Y_3	
	estimate	p-value	estimate	p-value	estimate	p-value
β_0	1.2221	0.0000*	4.8053	0.0000*	5.9204	0.0000*
β_1	0.0054	0.0000*	0.0005	0.6579	0.0002	0.8723
β_2	−0.0048	0.0354*	−0.0033	0.3210	−0.0059	0.1520
β_3	0.0065	0.0003*	−0.0053	0.0860	−0.0072	0.0493*
β_4	0.0015	0.0000*	0.0026	0.0000*	0.0044	0.0000*
β_5	0.0004	0.8572	−0.0013	0.7161	−0.0148	0.0003*
β_6	−0.0040	0.0122*	−0.0069	0.0036*	−0.0010	0.8617

Note: * indicates significance at $\alpha = 5\%$.

Table 9

Optimal spatial bandwidth for each period

Period	2017	2018	2019	2020	2021
Bandwidth	0.4338	0.5706	0.4779	0.5139	0.3496
Deviance	520.1691	555.3310	549.0274	544.4642	526.9951

Table 10

Deviance model based on temporal bandwidth

Temporal bandwidth	1	2	3	4	5
Deviance	2,703.302	2,581.551	2,594.740	2,602.725	2,587.411

Modelling educational indicators using GTWMGGR

This study proposes the GTWMGGR model, which is based on information on the distribution of response variables following the MGG distribution and accommodates spatio-temporal effects that impact the relationship mechanism of response variables and predictor variables. This model is applied with the aim of obtaining factors that affect MYS, SER and GER locally in each district/city in Central Java. Modelling is conducted using the geographical coordinates of the central location of the district/city and the period of observation in the parameter estimation. One important factor in the GTWMGGR model is the bandwidth used in forming the spatio-temporal weight matrix. In this study, determining the optimal bandwidth is conducted in two stages. The first stage is selecting the optimal spatial bandwidth in each observation period using the golden section search method. At this stage, the optimal bandwidth is determined using the GWMGGR model, which is an MGGR model that is weighted using a spatial weight matrix. The second stage is determining the optimal temporal weights of the GTWMGGR model using the grid search method based on the smallest deviance. The weight function used is a fixed Gaussian kernel function. The optimal spatial bandwidth for each observation period is presented in Table 9, and Table 10 presents the goodness of fit measure for the GTWMGGR model for each period to determine the best GTWMGGR model. Table 10 indicates that the optimal temporal bandwidth for the GTWMGGR model is two because it has the smallest deviance, showing that the average diversity of data up to two previous periods has a significant influence on the current period.

Table 11

Comparison of model accuracy

Model	Deviance	Log-likelihood
MGGR	2,811.262	−1,405.631
GWMGGR	2,606.462	−1,303.231
GTWMGGR	2,581.551	−1,290.776

Table 12

Model similarity test

No	Tested models	G^2	Df	F	p-value
1	MGGR	337.4331	18	2.8422	0.0012*
	GWMGGR	415.8942	63.0558		
2	MGGR	337.4331	18	2.7337	0.0017*
	GTWMGGR	436.5983	63.6676		
3	GWMGGR	415.8942	63.0558	0.9618	0.5612
	GTWMGGR	436.5983	63.6676		

Note: * indicates significance at $\alpha = 5\%$.

Table 11 presents the log-likelihood and deviance from MGGR, GWMGGR and GTWMGGR models. Deviance decreases in the model change from MGGR to GWMGGR, then from GWMGGR to GTWMGGR, indicating that GTWMGGR is the best model. The GTWMGGR model produces the smallest deviance because it accommodates spatio-temporal effects that impact the magnitude of the relationship between variables in addition to accommodating the relationship between response and predictor variables. This finding is reinforced by the results in Table 12 regarding the similarity between MGGR and GTWMGGR models. This test also indicates the significance level of the spatio-temporal effect in the GTWMGGR model. Based on Table 12, the F-test statistic value is 2.7337 with a p-value of 0.0017 (significant at the 1% level), indicating that the GTWMGGR model is significantly different from the MGGR model. Therefore, the GTWMGGR model is effective in accommodating spatio-temporal effects that significantly affect the magnitude of the relationships between response and predictor variables. However, the test in Table 12 row 3 indicates no significant difference between GTWMGGR and GWMGGR models. Therefore, it should be noted that in this case, temporal heterogeneity does not have a significant effect, while spatial heterogeneity has a highly dominant effect.

Table 13

Simultaneous test of GTWMGGR model parameters

Source	Deviance	G^2	Df	p-value
Null model	3,018.150	436.5983	63.6676	0.0000
GTWMGGR	2,581.551			

Table 14

Partial test of GTWMGGR parameters at the 33rd location (Purbalingga district)

Parameter	Y_1		Y_2		Y_3	
	estimate	p-value	estimate	p-value	estimate	p-value
β_0	1.1494	0.0000	4.4996	0.0000	5.2733	0.0000
β_1	0.0058	0.0000*	0.0005	0.5856	0.0003	0.7850
β_2	−0.0043	0.0212*	−0.0049	0.0272*	−0.0031	0.3340
β_3	0.0081	0.0000*	0.0001	0.9662	0.0010	0.7398
β_4	0.0015	0.0000*	0.0025	0.0000*	0.0054	0.0000*
β_5	−0.0010	0.5874	−0.0032	0.2822	−0.0169	0.0000*
β_6	−0.0039	0.0805	−0.0102	0.0000*	−0.0062	0.1656

Note: * indicates significance at $\alpha = 5\%$.

After determining the optimal model parameter estimator, we conduct statistical inference through hypothesis testing of the GTWMGGR model regression parameters and model interpretation. The results in Table 13 indicate that the null hypothesis (H_0) is rejected, suggesting that at least one predictor variable substantially affects the GTWMGGR model. Subsequently, we conduct partial parameter testing to determine the predictor variable that has a noteworthy impact on the response variable at each site. The partial test results for the GTWMGGR model parameters differ for each observation location due to the spatio-temporal weights used in the GTWMGGR model. For example, the estimation results and partial test of model parameters in Purbalingga district (33rd location) are presented in Table 14. The results indicate that at the 5% significance level, predictor variables X_1, X_2, X_3 and X_4 significantly affect response variable Y_1. Additionally, predictor variables X_2, X_4 and X_6 substantially affect response variable Y_2, while X_4 and X_5 have a major effect on response variable Y_3. Based on the partial test results at each observation location, we can group districts/cities that have similarities according to predictor variables that significantly affect the responses. Eleven groups are identified for MYS and GER indicators and nine groups for the SER indicators (Table 15). A surface map is presented in Figure 2 to illustrate the spread of each predictor's influence on the respective responses.

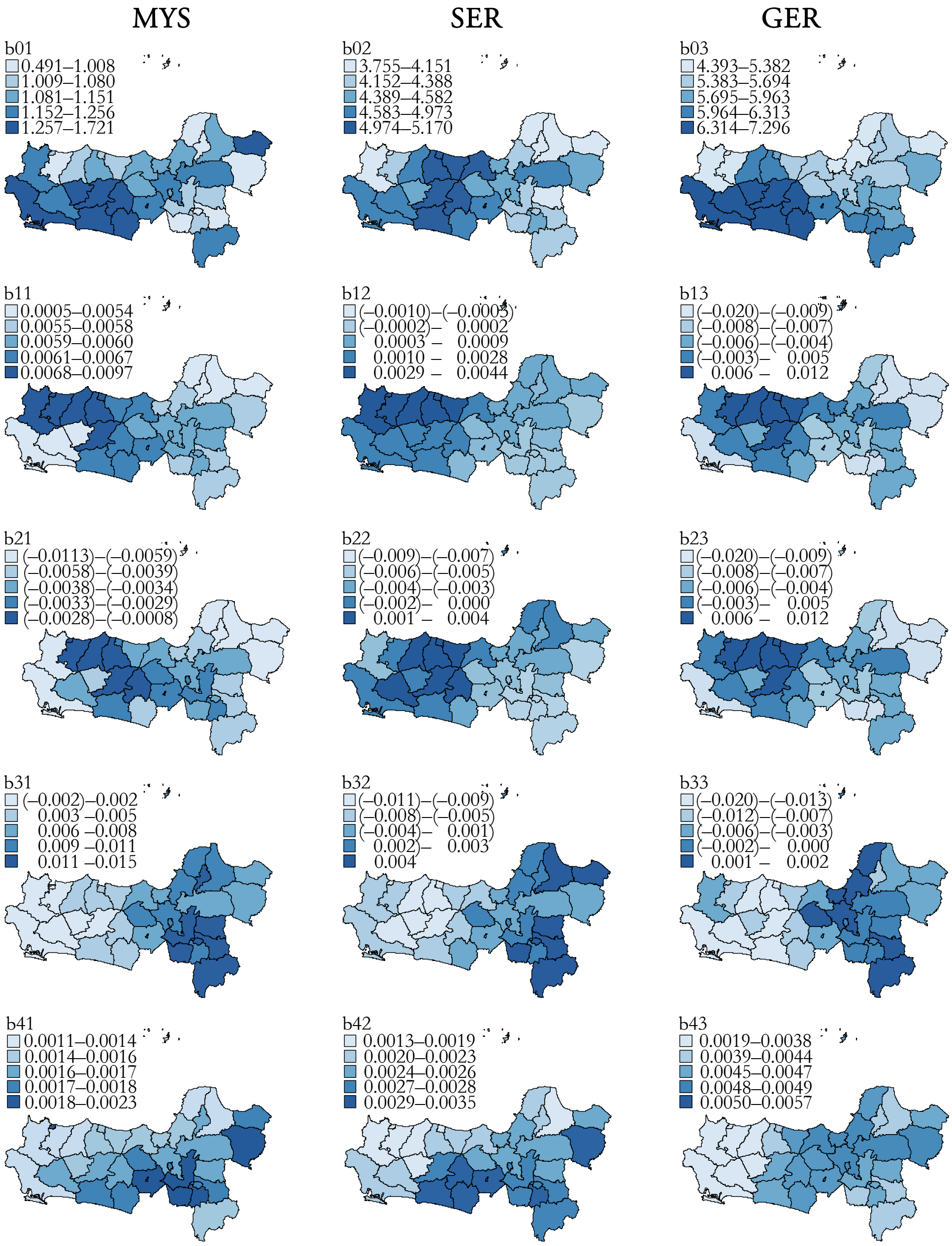
Our findings suggest that the achievement of MYS, GER and SER indicators in each district/city in Central Java is not only influenced by the predictors directly, but also by the inherent spatio-temporal heterogeneity effect. In 2023, Indonesia was not included in the top 20 countries and is in 67th position out of 203 countries. In addition, the average intelligence quotient of the Indonesian population is considered to be low, emphasising the need for comprehensive education reform to raise standards and improve global competitiveness [2]. Therefore, it is crucial to underline that improving education is not the exclusive responsibility of the government but requires active involvement from all relevant parties. An educational environment that supports the optimal development of children in Central Java can be developed through effective collaboration between the government, educational institutions, the business sector and the community, positively impacting the educational landscape and other aspects of regional development (Caraka et al. 2021, Kye et al. 2021).

Table 15

District/city clustering in Central Java based on the GTWMGGR model

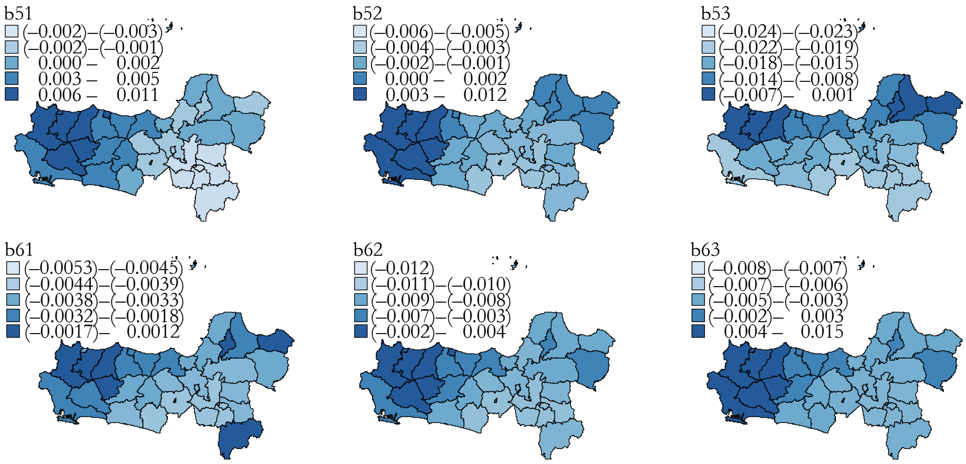
Response	Cluster	Location	Significant variables
MYS	1	1	X_1, X_2, X_4, X_5, and X_6
	2	2, 3	X_1, X_2, X_4, and X_5
	3	4	X_1, X_4, X_5, and X_6
	4	5, 6	X_1, X_4, and X_6
	5	7	X_1 and X_4
	6	8–11, 15, 22, 23, 30, 32	X_1, X_3, and X_4
	7	12, 14, 25, 27–29, 34	X_1, X_2, X_3, X_4, and X_5
	8	13, 26, 31	X_1, X_3, X_4, and X_5
	9	16–19, 35	X_1, X_2, X_3, X_4, X_5, and X_6
	10	20, 21	X_1, X_2, X_3, X_4, and X_6
	11	24, 33	X_1, X_2, X_3, and X_4
SER	1	1	X_3, X_4, X_5, and X_6
	2	2, 3, 28	X_1, X_3, X_4, and X_5
	3	4, 20, 24	X_1, X_3, X_4, and X_6
	4	5, 6, 8, 30	X_4 and X_6
	5	7	X_3, X_4, and X_6
	6	9–15, 21–23, 31–33	X_2, X_4, and X_6
	7	16–19, 35	X_1, X_2, X_3, X_4, X_5, and X_6
	8	25–27, 34	X_1, X_3, and X_4
	9	29	X_1, X_2, X_3, X_4, and X_5
GER	1	1, 3, 16–19, 35	X_1, X_2, X_3, X_4, X_5, and X_6
	2	2, 4	X_1, X_3, X_4, and X_5
	3	5, 6, 7	X_3, X_4, and X_5
	4	8, 12, 23, 30, 32	X_2, X_4, and X_5
	5	9–11, 13, 20, 31	X_2, X_4, X_5, and X_6
	6	14	X_4, X_5, and X_6
	7	15, 21, 22, 33	X_4 and X_5
	8	24	X_1, X_4, and X_5
	9	25–27, 34	X_1, X_2, X_3, X_4, and X_5
	10	28	X_1, X_2, X_3, and X_4
	11	29	X_1, X_3, X_4, and X_6

Figure 2
Surface of the parameter estimates with the GTWMGGR model



(Figures continue on the next page.)

(Continued.)



Conclusion

This study proposes the GTWMGGR model to accommodate the effect of spatio-temporal heterogeneity in the MGGR model. Parameter estimation is conducted by applying the BHHH algorithm to optimise parameters in the MLE method. Model inference is conducted with three hypothesis tests, covering overall parameter testing, partial parameter testing and spatio-temporal model fit testing, determining the test statistic of these hypotheses based on the likelihood ratio test. The results of simulation and empirical studies demonstrate that the GTWMGGR model is more powerful than the MGGR model for data significantly affected by spatio-temporal heterogeneity. This model implies that the significance of predictor variables on each response in each location is unique because this model produces localised parameter estimates. Spatial clustering of each response variable based on predictors' significance is a novel approach for understanding the mechanism of the relationships between response and predictor variables in a spatial framework to create specifically targeted interventions for each site.

However, some notes related to the results of this study must be addressed. The hat matrix **S** inadequately characterises the effective count of local shape and scale parameters. As a result, the estimation of the effective number of parameters in GTWMGGR may be underestimated due to the difficulty of accurately determining this number. Therefore, it will be interesting to investigate this issue in more depth in our future research. Parameter estimation using the GWMGGR model is also time-consuming; therefore, it is recommended to further optimise the likelihood function by using a more computationally efficient approaches based on non-gradient methods. We will also consider the use of Monte Carlo permutation and bootstrap tests for spatial or spatio-temporal non-stationarity detection as in general GWR models. The novelty of this test is its ability to detect which predictors have local and global effects, leading to the development of a mixed model that incorporates both variables. The final issue is about mapping GTWMGGR results. A better approach may be to visualise the estimated coefficients using a three-dimensional mapping technique combined with simultaneous significance testing for the equality of each parameter across all dependent variables; however, such simultaneous hypothesis testing is not available in this study. In addition, no technique or software is available to perform three-dimensional mapping. Therefore, future research should address these issues.

Acknowledgement

The first author thanks the financial support provided by Balai Pembiayaan Pendidikan Tinggi (BPPT) or Central for Higher Education Funding under the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) and Lembaga Pengelola Dana Pendidikan (LPDP) under the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia.

Appendix

A1. The gradient of GTWMGGR parameter $\mathbf{g}_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)$

$$\mathbf{g}_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)=\left[\frac{\partial L_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)}{\partial \boldsymbol{\beta}_{1i*}^{T}}\cdots\frac{\partial L_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)}{\partial \boldsymbol{\beta}_{Ki*}^{T}}\frac{\partial L_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)}{\partial \lambda_{i*}}\frac{\partial L_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)}{\partial \tau_{i*}}\frac{\partial L_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)}{\partial \boldsymbol{\delta}_{i*}^{T}}\right]^{T}$$

$$\mathbf{g}_{il}\left(\mathbf{\Theta}_{GTR}^{i*}\right)=w_{ii*}^{l}\left[\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \boldsymbol{\beta}_{1i*}^{T}}\cdots\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \boldsymbol{\beta}_{Ki*}^{T}}\right.$$

$$\left.\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \lambda_{i*}}\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \tau_{i*}}\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \boldsymbol{\delta}_{i*}^{T}}\right]^{T}$$

– The first partial derivative of $\log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)$ with respect to $\boldsymbol{\beta}_{ki*}^{T}$

For $k=1$,

$$\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \boldsymbol{\beta}_{1i*}^{T}}=-\tau_{i*}\lambda_{i*}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{i}^{T}\boldsymbol{\beta}_{1i*}\right)}{\exp\left(\mathbf{x}_{i}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}}\right)+\tau_{i*}\lambda_{i*}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{2i*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{2i*}}\right)$$

$$+\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)\left(y_{1il}-\delta_{1i*}\right)^{\tau_{i*}}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}\right)^{\tau_{i*}+1}}$$

$$-\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)\left(y_{2il}-y_{1il}-\delta_{2i*}\right)^{\tau_{i*}}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{2i*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{2i*}\right)^{\tau_{i*}+1}}$$

For $k=2,\cdots,K-1$,

$$\frac{\partial \log f\left(\mathbf{y}_{il}\left|\mathbf{\Theta}_{GTR}^{i*}\right.\right)}{\partial \boldsymbol{\beta}_{ki*}^{T}}=-\tau_{i*}\lambda_{i*}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}}\right)$$

$$+\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)\left(y_{kil}-y_{(k-1)il}-\delta_{ki*}\right)^{\tau_{i*}}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}\right)^{\tau_{i*}+1}}\right)$$

$$+\tau_{i*}\lambda_{i*}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k+1)i*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\delta_{(k+1)i*}}\right)$$

$$-\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)\left(y_{(k+1)il}-y_{kil}-\delta_{(k+1)i*}\right)^{\tau_{i*}}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k+1)i*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\delta_{(k+1)i*}\right)^{\tau_{i*}+1}}\right)$$

For $k = K$,

$$\frac{\partial\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)}{\partial\boldsymbol{\beta}_{Ki*}^{T}}=-\tau_{i*}\lambda_{i*}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{Ki*}\right)}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{Ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(K-1)i*}\right)-\delta_{Ki*}}\right)$$

$$+\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\left(\frac{\mathbf{x}_{il}\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{Ki*}\right)\left(y_{Kil}-y_{(K-1)il}-\delta_{Ki*}\right)^{\tau_{i*}}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{Ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(K-1)i*}\right)-\delta_{Ki*}\right)^{\tau_{i*}+1}}\right)$$

– The first partial derivative of $\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)$ with respect to λ_{i*}

$$\frac{\partial\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)}{\partial\lambda_{i*}}=\frac{\partial A_{1\lambda}}{\partial\lambda_{i*}}+\sum_{k=2}^{K}\frac{\partial A_{k\lambda}}{\partial\lambda_{i*}}\text{, with}$$

$$\frac{\partial A_{1\lambda}}{\partial\lambda_{i*}}=-\left(\tau_{i*}\log\left(\Gamma\left(\lambda_{i*}\right)\right)+\left(\tau_{i*}\lambda_{i*}+1\right)\Psi\left(\lambda_{i*}\right)\right)\quad+\tau_{i*}\left(\log\left(\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)+\lambda_{i*}\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)$$

$$+\tau_{i*}\log\left(y_{1il}-\delta_{1i*}\right)-\tau_{i*}\log\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}\right)$$

$$+\tau\left(\frac{\Gamma\left(\lambda+\frac{1}{\tau}\right)}{\Gamma\left(\lambda\right)}\right)^{\tau}\left(\Psi\left(\lambda\right)-\Psi\left(\lambda+\frac{1}{\tau}\right)\right)\left(\frac{y_{1il}-\delta_{1}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1}\right)-\delta_{1}}\right)^{\tau}$$

$$\frac{\partial A_{k\lambda}}{\partial\lambda_{i*}}=-\left(\tau_{i*}\log\left(\Gamma\left(\lambda_{i*}\right)\right)+\left(\tau_{i*}\lambda_{i*}+1\right)\Psi\left(\lambda_{i*}\right)\right)+\tau_{i*}\left(\log\left(\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)+\lambda_{i*}\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)$$

$$+\tau_{i*}\log\left(y_{kil}-y_{(k-1)il}-\delta_{k}\right)-\tau_{i*}\log\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{k}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{k-1}\right)-\delta_{k}\right)$$

$$+\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma(\lambda_{i*})}\right)^{\tau_{i*}}\left(\Psi(\lambda_{i*})-\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)\left(\frac{y_{kil}-y_{(k-1)il}-\delta_{ki*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}}\right)^{\tau_{i*}}$$

$\Psi(\bullet)$ is the digamma function.

– The first partial derivative of $\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)$ with respect to τ_{i*}

$$\frac{\partial \log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)}{\partial \tau_{i*}}=\frac{\partial A_{1\tau}}{\partial \tau_{i*}}+\sum_{k=2}^{K}\frac{\partial A_{k\tau}}{\partial \tau_{i*}}\text{, with}$$

$$\frac{\partial A_{1\tau}}{\partial \tau_{i*}}=\frac{1}{\tau_{i*}}-\lambda_{i*}\log\left(\Gamma(\lambda_{i*})\right)+\lambda_{i*}\log\left(\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)-\frac{\lambda_{i*}}{\tau_{i*}}\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)$$
$$+\lambda_{i*}\log\left(y_{1il}-\delta_{1i*}\right)-\lambda_{i*}\log\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}\right)$$
$$-\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma(\lambda_{i*})}\right)^{\tau_{i*}}\left(\log\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma(\lambda_{i*})}\right)-\frac{\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\tau_{i*}}\right)\left(\frac{y_{1il}-\delta_{1i*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}}\right)^{\tau_{i*}}$$
$$-\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma(\lambda_{i*})}\right)^{\tau_{i*}}\left(\frac{y_{1il}-\delta_{1i*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1}\right)-\delta_{1i*}}\right)^{\tau_{i*}}\log\left(\frac{y_{1il}-\delta_{1i*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}}\right)$$

$$\frac{\partial A_{k\tau}}{\partial \tau_{i*}}=\frac{1}{\tau_{i*}}-\lambda_{i*}\log\left(\Gamma(\lambda_{i*})\right)+\lambda_{i*}\log\left(\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)\right)-\frac{\lambda_{i*}}{\tau_{i*}}\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)$$
$$+\lambda_{i*}\log\left(y_{kil}-y_{(k-1)il}-\delta_{ki*}\right)-\lambda_{i*}\log\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}\right)$$
$$-\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma(\lambda_{i*})}\right)^{\tau_{i*}}\left(\ln\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma(\lambda_{i*})}\right)-\frac{\Psi\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\tau_{i*}}\right)$$

$$\left(\frac{y_{kil}-y_{(k-1)il}-\delta_{ki*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}}\right)^{\tau_{i*}}-\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}$$

$$\left(\frac{y_{kil}-y_{(k-1)il}-\delta_{ki*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}}\right)^{\tau_{i*}}\log\left(\frac{y_{kil}-y_{(k-1)il}-\delta_{ki*}}{\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}}\right)$$

– The first partial derivative of $\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)$ with respect to $\delta_{ki*};k=1,2,\cdots,K$

For $k=1$,

$$\frac{\partial\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)}{\partial\delta_{1i*}}=\left(1-\tau_{i*}\lambda_{i*}\right)\left(y_{1il}-\delta_{1}\right)^{-1}+\tau_{i*}\lambda_{i*}\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}\right)^{-1}$$

$$+\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\frac{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-y_{1il}\right)\left(y_{1il}-\delta_{1i*}\right)^{\tau_{i*}-1}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{1i*}\right)-\delta_{1i*}\right)^{\tau_{i*}+1}}$$

For $k=2,3,\cdots,K$,

$$\frac{\partial\log f\left(\mathbf{y}_{il}\left|\boldsymbol{\Theta}_{GTR}^{i*}\right.\right)}{\partial\delta_{1i*}}=\left(1-\tau_{i*}\lambda_{i*}\right)\left(y_{kil}-y_{(k-1)il}-\delta_{ki*}\right)^{-1}+\tau_{i*}\lambda_{i*}\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}\right)^{-1}$$

$$+\tau_{i*}\left(\frac{\Gamma\left(\lambda_{i*}+\frac{1}{\tau_{i*}}\right)}{\Gamma\left(\lambda_{i*}\right)}\right)^{\tau_{i*}}\frac{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-y_{kil}+y_{(k-1)il}\right)\left(y_{kil}-y_{(k-1)il}-\delta_{ki*}\right)^{\tau_{i*}-1}}{\left(\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{ki*}\right)-\exp\left(\mathbf{x}_{il}^{T}\boldsymbol{\beta}_{(k-1)i*}\right)-\delta_{ki*}\right)^{\tau_{i*}+1}}$$

A2. Distribution of the statistical test G_{GTR}^{2}

$$G_{GTR}^{2}=2\left(L\left(\hat{\Omega}_{GTR}\right)-L\left(\hat{\omega}_{GTR}\right)\right)\xrightarrow[n\to\infty]{d}\chi_{K\text{tr}(\mathbf{S})}^{2}.$$

If $L\left(\hat{\Omega}_{GTR}\right)$ and $L\left(\hat{\omega}_{GTR}\right)$ represents the GTWMGGR log-likelihood under population and log-likelihood H_0. So,

$$G_{GTR}^{2}=2\left(L\left(\hat{\Omega}_{GTR}\right)-L\left(\omega_{GTR}\right)\right)-2\left(L\left(\hat{\omega}_{GTR}\right)-L\left(\omega_{GTR}\right)\right)$$

$L\left(\omega_{GTR}\right)$ function can be approached by Taylor's second-degree expansion around $\hat{\Omega}_{GTR}$ as follows.

$$L(\omega_{GTR}) \approx L(\hat{\Omega}_{GTR}) + \mathbf{g}(\hat{\Omega}_{GTR})(\omega_{GTR} - \hat{\Omega}_{GTR}) \quad -\frac{1}{2}(\omega_{GTR} - \hat{\Omega}_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\omega_{GTR} - \hat{\Omega}_{GTR})$$

where $\mathbf{g}(\hat{\Omega}_{GTR}) = \left.\frac{\partial \ln L(\Omega_{GTR})}{\partial(\Omega_{GTR})}\right|_{\Omega_{GTR}=\hat{\Omega}_{GTR}} = \mathbf{0}$ and

$$\mathbf{I}(\hat{\Omega}_{GTR}) = -\left.\frac{\partial^2 \ln L(\Omega_{GTR})}{\partial(\Omega_{GTR})\partial(\Omega_{GTR})^T}\right|_{\Omega_{GTR}=\hat{\Omega}_{GTR}}.$$

Because $\mathbf{g}(\hat{\Omega}_{GTR}) = \mathbf{0}$, then

$$L(\omega_{GTR}) \approx L(\hat{\Omega}_{GTR}) - \frac{1}{2}(\omega_{GTR} - \hat{\Omega}_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\omega_{GTR} - \hat{\Omega}_{GTR})$$

$$2\left(L(\hat{\Omega}_{GTR}) - L(\omega_{GTR})\right) \approx (\hat{\Omega}_{GTR} - \omega_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\hat{\Omega}_{GTR} - \omega_{GTR})$$

Then, $L(\omega_{GTR})$ function also approached by Taylor's second-degree expansion around $\hat{\omega}_{GTR}$ as follows.

$$L(\omega_{GTR}) \approx L(\hat{\omega}_{GTR}) + \mathbf{g}(\hat{\Omega}_{GTR})(\omega_{GTR} - \hat{\omega}_{GTR}) \quad -\frac{1}{2}(\omega_{GTR} - \hat{\omega}_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\omega_{GTR} - \hat{\omega}_{GTR})$$

Because $\mathbf{g}(\hat{\Omega}_{GTR}) = \mathbf{0}$, then

$$L(\omega_{GTR}) \approx L(\hat{\omega}_{GTR}) - \frac{1}{2}(\omega_{GTR} - \hat{\omega}_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\omega_{GTR} - \hat{\omega}_{GTR})$$

$$2\left(L(\hat{\omega}_{GTR}) - L(\omega_{GTR})\right) \approx (\hat{\omega}_{GTR} - \omega_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\hat{\omega}_{GTR} - \omega_{GTR})$$

Therefore, G^2_{GTR} can be written as

$$G^2_{GTR} = 2\left(L(\hat{\Omega}_{GTR}) - L(\omega_{GTR})\right) - 2\left(L(\hat{\omega}_{GTR}) - L(\omega_{GTR})\right)$$

$$G^2_{GTR} \approx (\hat{\Omega}_{GTR} - \omega_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\hat{\Omega}_{GTR} - \omega_{GTR}) - (\hat{\omega}_{GTR} - \omega_{GTR})^T \left[\mathbf{I}(\hat{\Omega}_{GTR})\right](\hat{\omega}_{GTR} - \omega_{GTR})$$

$\hat{\Omega}_{GTR}$ and $\hat{\omega}_{GTR}$ are estimated by the MLE method, so we can partition the two parameter vectors as follows.

$$\hat{\Omega}_{GTR} = \left[\hat{\boldsymbol{\beta}}^{*T}_{GTR} \quad \hat{\boldsymbol{\Theta}}^{*T}_{1GTR}\right]^T$$

where $\hat{\boldsymbol{\beta}}^{*}_{GTR} = \left[\hat{\boldsymbol{\beta}}^{*T}_{11} \ \hat{\boldsymbol{\beta}}^{*T}_{21} \cdots \hat{\boldsymbol{\beta}}^{*T}_{K1} \cdots \hat{\boldsymbol{\beta}}^{*T}_{12} \ \hat{\boldsymbol{\beta}}^{*T}_{22} \cdots \hat{\boldsymbol{\beta}}^{*T}_{K2} \cdots \hat{\boldsymbol{\beta}}^{*T}_{1n} \ \hat{\boldsymbol{\beta}}^{*T}_{2n} \cdots \hat{\boldsymbol{\beta}}^{*T}_{Kn}\right]^T$,

$\hat{\mathbf{\Theta}}_{1GTR}^{*} = \left[\hat{\boldsymbol{\beta}}_{0GTR}^{T} \ \hat{\lambda}_1 \ \hat{\lambda}_2 \cdots \hat{\lambda}_n \ \hat{\tau}_1 \ \hat{\tau}_2 \cdots \hat{\tau}_n \ \hat{\boldsymbol{\delta}}_1^{T} \ \hat{\boldsymbol{\delta}}_2^{T} \cdots \hat{\boldsymbol{\delta}}_n^{T} \right]^{T}$ and

$\hat{\boldsymbol{\beta}}_{0GTR} = \left[\hat{\beta}_{011} \ \hat{\beta}_{021} \cdots \hat{\beta}_{0K1} \cdots \hat{\beta}_{012} \ \hat{\beta}_{022} \cdots \hat{\beta}_{0K2L} \cdots \hat{\beta}_{01n} \ \hat{\beta}_{02n} \cdots \hat{\beta}_{0Kn} \right]^{T}$.

$\hat{\omega}_{GTR} = \left[\mathbf{0}_{(pKn \times 1)} \quad \hat{\mathbf{\Theta}}_{\omega 1GTR}^{*T} \right]^{T}$

where $\hat{\mathbf{\Theta}}_{\omega 1GTR}^{*} = \left[\hat{\boldsymbol{\beta}}_{\omega 0GTR}^{T} \ \hat{\lambda}_{\omega 1} \ \hat{\lambda}_{\omega 2} \cdots \hat{\lambda}_{\omega n} \ \hat{\tau}_{\omega 1} \ \hat{\tau}_{\omega 2} \cdots \hat{\tau}_{\omega n} \ \hat{\boldsymbol{\delta}}_{\omega 1}^{T} \ \hat{\boldsymbol{\delta}}_{\omega 2}^{T} \cdots \hat{\boldsymbol{\delta}}_{\omega n}^{T} \right]^{T}$ and

$\hat{\boldsymbol{\beta}}_{\omega 0GTR} = \left[\hat{\beta}_{\omega 011} \ \hat{\beta}_{\omega 021} \cdots \hat{\beta}_{\omega 0K1} \cdots \hat{\beta}_{\omega 0ki} \cdots \hat{\beta}_{\omega 01n} \ \hat{\beta}_{\omega 02n} \cdots \hat{\beta}_{\omega 0Kn} \right]^{T}$.

Using block matrix theory, the quadratic form $\left(\hat{\omega}_{GTR} - \omega_{GTR} \right)^{T} \left[\mathbf{I}\left(\hat{\Omega}_{GTR} \right) \right] \left(\hat{\omega}_{GTR} - \omega_{GTR} \right)$ can be written as follows.

$$\left(\hat{\omega}_{GTR} - \omega_{GTR} \right)^{T} \left[\mathbf{I}\left(\hat{\Omega}_{GTR} \right) \right] \left(\hat{\omega}_{GTR} - \omega_{GTR} \right) = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{GTR}^{*} \\ \hat{\mathbf{\Theta}}_{1GTR}^{*} - \mathbf{\Theta}_{\omega 1GTR}^{*} \end{bmatrix}^{T} \begin{bmatrix} [\mathbf{I}_{12}][\mathbf{I}_{22}]^{-1}[\mathbf{I}_{21}] & [\mathbf{I}_{12}] \\ [\mathbf{I}_{21}] & [\mathbf{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{GTR}^{*} \\ \hat{\mathbf{\Theta}}_{1GTR}^{*} - \mathbf{\Theta}_{\omega 1GTR}^{*} \end{bmatrix}.$$

The quadratic form $\left(\hat{\Omega}_{GTR} - \omega_{GTR} \right)^{T} \left[\mathbf{I}\left(\hat{\Omega}_{GTR} \right) \right] \left(\hat{\Omega}_{GTR} - \omega_{GTR} \right)$ can be written as follows.

$$\left(\hat{\Omega}_{GTR} - \omega_{GTR} \right)^{T} \left[\mathbf{I}\left(\hat{\Omega}_{GTR} \right) \right] \left(\hat{\Omega}_{GTR} - \omega_{GTR} \right) = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{GTR}^{*} \\ \hat{\mathbf{\Theta}}_{1GTR}^{*} - \mathbf{\Theta}_{\omega 1GTR}^{*} \end{bmatrix}^{T} \begin{bmatrix} [\mathbf{I}_{11}] & [\mathbf{I}_{12}] \\ [\mathbf{I}_{21}] & [\mathbf{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{GTR}^{*} \\ \hat{\mathbf{\Theta}}_{1GTR}^{*} - \mathbf{\Theta}_{\omega 1GTR}^{*} \end{bmatrix}.$$

Therefore, the statistical test G_{GTR}^{2} can be simplified to

$$G_{GTR}^{2} \approx \hat{\boldsymbol{\beta}}_{GTR}^{*T} [\mathbf{I}_{11}] \hat{\boldsymbol{\beta}}_{GTR}^{*} - \hat{\boldsymbol{\beta}}_{GTR}^{*T} [\mathbf{I}_{12}][\mathbf{I}_{22}]^{-1}[\mathbf{I}_{21}] \hat{\boldsymbol{\beta}}_{GTR}^{*} \approx \hat{\boldsymbol{\beta}}_{GTR}^{*T} [\mathbf{I}_{11}]^{-1} \hat{\boldsymbol{\beta}}_{GTR}^{*}.$$

So, this can be obtained:

$$\hat{\boldsymbol{\beta}}_{GTR}^{*} \xrightarrow[n \to \infty]{d} N\left(\mathbf{0}, \left[\mathbf{I}^{11} \right]_{(pKn \times pKn)} \right), \text{ and } \left[\mathbf{I}^{11} \right]^{-\frac{1}{2}} \hat{\boldsymbol{\beta}}_{GTR}^{*} \xrightarrow[n \to \infty]{d} N\left(\mathbf{0}, \mathbf{I}_{pKn} \right).$$

Thus, it is obtained that:

$$G_{GTR}^{2} = 2\left(L\left(\hat{\Omega}_{GTR} \right) - L\left(\hat{\omega}_{GTR} \right) \right) \approx \left[\left[\mathbf{I}^{11} \right]^{-\frac{1}{2}} \hat{\boldsymbol{\beta}}_{GTR}^{*} \right]^{T} \left[\left[\mathbf{I}^{11} \right]^{-\frac{1}{2}} \hat{\boldsymbol{\beta}}_{GTR}^{*} \right]$$

$$= \mathbf{z}^{T} \mathbf{z} \xrightarrow[n \to \infty]{d} \chi_{pKn}^{2}$$

where $\mathbf{z} = \left[\mathbf{I}^{11} \right]^{-\frac{1}{2}} \hat{\boldsymbol{\beta}}_{GTR}^{*} \xrightarrow[n \to \infty]{d} N\left(\mathbf{0}, \mathbf{I}_{pKn} \right)$.

pKn is the number of elements of vector $\boldsymbol{\beta}^{*}_{GTR}$. However, according to the determination of the degrees of freedom in the GWR model by Fotheringham et al. (2002), df_{GTR} can be approximated by the effective number of parameters in the GTWMGGR model, which is defined as

$$df_{GTR} = \sum_{k=1}^{K} \operatorname{tr}(\mathbf{S}) = K \operatorname{tr}(\mathbf{S})$$

where $\mathbf{S}$ express the GTWMGGR projection matrix which written in equation (22).

Accordingly, it can be concluded that $G^2_{GTR} \xrightarrow[n\to\infty]{d} \chi^2_{K\operatorname{tr}(\mathbf{S})}$. ■

A3. Distribution of the statistical test Z_{jki}

$$Z_{jki} = \frac{\hat{\beta}_{jki}}{\sqrt{\widehat{\operatorname{var}}\left(\hat{\beta}_{jki}\right)}} \xrightarrow[n\to\infty]{d} N(0,1)$$

If $L\left(\hat{\Omega}^i_{GTR}\right)$ and $L\left(\hat{\omega}^i_{GTR}\right)$ represents the GTWMGGR log-likelihood under population and log-likelihood under H_0 to test the significance of parameter partially at *i*-th location.

$\hat{\Omega}^i_{GTR}$ and $\hat{\omega}^i_{GTR}$ are estimated using MLE method, so we can partition the two parameter vectors as follows.

$$\hat{\Omega}^i_{GTR} = \left[\hat{\beta}_{jki} \quad \left[\hat{\boldsymbol{\Theta}}^i_{2GTR}\right]^T\right]^T,$$

where $\hat{\boldsymbol{\Theta}}^i_{2GTR} = \left[\hat{\beta}_{01i} \ \hat{\beta}_{11i} \cdots \hat{\beta}_{(j-1)ki} \ \hat{\beta}_{(j+1)ki} \cdots \hat{\beta}_{pKi} \ \hat{\lambda}_i \ \hat{\tau}_i \ \hat{\boldsymbol{\delta}}^T_i\right]^T$, and

$$\hat{\omega}^i_{GTR} = \left[0 \quad \left[\hat{\boldsymbol{\Theta}}^i_{\omega GTR}\right]^T\right]^T$$

where $\hat{\boldsymbol{\Theta}}^i_{\omega GTR} = \left[\hat{\beta}_{\omega 0i} \ \hat{\beta}_{\omega 11i} \cdots \hat{\beta}_{\omega(j-1)ki} \ \hat{\beta}_{\omega(j+1)ki} \cdots \hat{\beta}_{\omega pKi} \ \hat{\lambda}_{\omega i} \ \hat{\tau}_{\omega i} \ \hat{\boldsymbol{\delta}}^T_{\omega i}\right]^T$.

Therefore, we can obtain that:

$$\begin{aligned}
\chi^2_{GTR} &\approx \left(\hat{\beta}_{jki} - 0\right)^T \left([\mathbf{I}_{11}] - [\mathbf{I}_{12}][\mathbf{I}_{22}]^{-1}[\mathbf{I}_{21}]\right)\left(\hat{\beta}_{jki} - 0\right) \\
&\approx \left(\hat{\beta}_{jki} - 0\right)^T \left[\mathbf{I}^{11}\right]^{-1}\left(\hat{\beta}_{jki} - 0\right) \\
&\approx \hat{\beta}^2_{jki}\left[\mathbf{I}^{11}\right]^{-1}.
\end{aligned}$$

Based on the asymptotic normality of the MLE estimator, it is known that:

$$\sqrt{nL}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}-\mathrm{E}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right)\xrightarrow[n\to\infty]{d}N\left(\mathbf{0},\left[\mathbf{I}\left(\mathrm{E}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right)\right]^{-1}\right),$$

where $\mathbf{I}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)$ is *Fisher Information* matrix for GTWMGGR parameters at *i*-th location. Hogg et al. (2019) they stated that:

$$\mathbf{I}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)=\mathrm{cov}\left[\mathbf{g}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right]=-\mathrm{E}\left(\mathbf{g}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\left[\mathbf{g}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right]^{T}\right)=-\mathrm{E}\left[\mathbf{H}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right], \text{ and}$$

$$\left[\mathbf{I}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right]^{-1}=\left[-\mathrm{E}\left[\mathbf{H}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right]\right]^{-1}.$$

So, we can conclude that:

$$\chi^2_{GTR}\approx\hat{\beta}^2_{jki}\left[\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)\right]^{-1}=\frac{\hat{\beta}^2_{jki}}{\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)}\xrightarrow[n\to\infty]{d}\chi^2_1,$$

or, using the square root of χ^2_{GTR}, we can obtain the following test statistic.

$$Z_{jki}=\frac{\hat{\beta}_{jki}}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)}}\xrightarrow[n\to\infty]{d}N(0,1),$$

$\widehat{\mathrm{Var}}\left(\hat{\beta}_{jki}\right)$ is the main diagonal element corresponding to $\hat{\beta}_{jki}$ of the matrix $-\left[\mathbf{H}\left(\hat{\boldsymbol{\Theta}}^{i}_{GTR}\right)\right]^{-1}$.

REFERENCES

BARTLETT, M. S. (1951): The effect of standardization on a X2 approximation in factor analysis *Biometrika* 38 (3/4): 337–344. http://doi.org/10.2307/2332580

BERNDT, E. K.–HALL, B. H.–HALL, R. E.–HAUSMAN, J. A. (1974): Estimation and inference in nonlinear structural models *Annals of Economic and Social Measurement* 3 (4): 653–665.

BERTUS, Z. (2017): Investigating the background of radical right-wing mobilization in Hungary with regional statistical methods *Regional Statistics* 7 (2): 190–208. https://doi.org/10.15196/RS070207

CARAKA, R. E.–NOH, M.–CHEN, R. C.–LEE, Y.–GIO, P. U.–PARDAMEAN, B. (2021): Connecting climate and communicable disease to Penta Helix using hierarchical likelihood structural equation modelling *Symmetry* 13(657): 1–21. https://doi.org/10.3390/sym13040657

CHEN, V. Y. J.–YANG, T. C.–JIAN, H. L. (2022): Geographically weighted regression modeling for multiple outcomes *Annals of the American Association of Geographers* 112 (5): 1278–1295. https://doi.org/10.1080/24694452.2021.1985955

CONOVER, W. J. (1971): *Practical nonparametric statistics* John Wiley & Sons., New York.

DA SILVA, A. R.–RODRIGUES, T. C. V. (2014): Geographically weighted negative binomial regression-incorporating overdispersion *Statistics and Computing* 24: 769–783. https://doi.org/10.1007/s11222-013-9401-9

DA SILVA, A. R.–De OLIVEIRA LIMA, A. (2017): Geographically weighted Beta regression *Spatial Statistics* 21: 279–303. https://doi.org/10.1016/j.spasta.2017.07.011

DIANTINI, N. L. S.–PURHADI-CHOIRUDDIN, A. (2023): Parameter estimation and hypothesis testing on three parameters log normal regression *AIP Conference Proceedings* 2554 (1): 030024. https://doi.org/10.1063/5.0104443

DU, Z.–WU, S.–ZHANG, F.–LIU, R.–ZHOU, Y. (2018): Extending geographically and temporally weighted regression to account for both spatiotemporal heterogeneity and seasonal variations in coastal seas *Ecological Informatics* 43: 185–199. https://doi.org/10.1016/j.ecoinf.2017.12.005

FÁBIÁN, Z. (2014): Method of the geographically weighted regression and an example for its application *Regional Statistics* 4 (1): 61–75. https://doi.org/10.15196/RS04105

FOTHERINGHAM, A. S.–BRUNSDON, C.–CHARLTON, M. (2002): *Geographically weighted regression: the analysis of spatially varying relationships* John Wiley & Sons Inc., Chichester.

FOTHERINGHAM, A. S.–CRESPO, R.–YAO, J. (2015): Geographical and temporal weighted regression (GTWR) *Geographical Analysis* 47 (4): 431–452. http://doi.org/10.1111/gean.12071

FOTHERINGHAM, A. S.–YANG, W.–KANG, W. (2017): Multiscale geographically weighted regression (MGWR) *Annals of the American Association of Geographers* 107 (6): 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

GREENE, W. H. (2008): *Econometric analysis (6th ed.)* Prentice Hall, New Jersey.

HARINI, S.–PURHADI–MASHURI, M.–SUNARYO, S. (2012): Statistical test for multivariate geographically weighted regression model using the method of maximum likelihood ratio test *International Journal of Applied Mathematics and Statistics* 29 (5): 110–115.

HOGG, R. V.–MCKEAN, J. W.–CRAIG, A. T. (2019): *Introduction to mathematical statistics* 8 Pearson, Boston.

HUANG, B.–WU, B.–BARRY, M. (2010): Geographically and temporally weighted regression for modelling spatiotemporal variation in house prices *International Journal of Geographical Information Science* 24 (3): 383–401. http://doi.org/10.1080/13658810802672469

KYE, B.−HAN, N.−KIM, E.−PARK, Y.−JO, S. (2021): Educational applications of metaverse: possibilities and limitations *Journal of Educational Evaluation for Health Professions* 18: 32. https://doi.org/10.3352/jeehp.2021.18.32

LIU, J.–ZHAO, Y.–YANG, Y.–XU, S.–ZHANG, F.–ZHANG, X.–SHI, L.–OIU, A. (2017): A mixed geographically and temporally weighted regression: exploring spatial-temporal variations from global and local perspectives *Entropy* 19 (2): 53. https://doi.org/10.3390/e19020053

PALMÍ-PERALES, F.–GÓMEZ-RUBIO, V.–BIVAND, R. S.–CAMELETTI, M.–RUE, H. (2023): Bayesian inference for multivariate spatial models with INLA *The R Journal* 15 (3):172–190. https://doi.org/10.32614/RJ-2023-068

PUTRI, D. E.–PURHADI–PRASTYO, D. D. (2017): Parameter estimation and hypothesis testing on geographically weighted gamma regression *Journal of Physics: Conference Series* 893: 012025. https://doi.org/10.1088/1742-6596/893/1/012025

RAHAYU, A.–PURHADI–SUTIKNO–PRASTYO, D. D. (2020): Multivariate gamma regression: parameter estimation, hypothesis testing, and its application *Symmetry* 12 (5): 813. https://doi.org/10.3390/sym12050813

SAFARALIZADEH, E.–JANAKIPOUR, F.–NIKKHOO, S.–AMIMI, M.–ZARGHAMFARD, M. (2024): Factors affecting housing prices in Metropolitan regions: the case of Tehran, 2021 *Regional Statistics* 14 (1): 130–158. https://doi.org/10.15196/RS140107

SANCHEZ, R.–MACKENZIE, S. A. (2016): Information thermodynamics of cytosine DNA methylation *PLoS ONE* 11 (3): 1–20. http://doi.org/10.1371/journal.pone.0150427

SHANKER, R.–SHUKLA, K. K. (2016): On modeling of lifetime data using three-parameter generalized lindley and generalized gamma distributions *Biometrics & Biostatistics International Journal* 4 (7): 283–288. http://doi.org/10.15406/bbij.2016.04.00117

SIFRIYANI, S.–RASJID, M.–ROSADI, D.–ANWAR, S.–WAHYUNI, R. D.–JALALUDDIN, S. (2022): Spatial-temporal epidemiology of Covid-19 using a geographically and temporally weighted regression model *Symmetry* 14 (4): 742. https://doi.org/10.3390/sym14040742

STACY, E. W. (1962): A Generalization of the gamma distribution *The Annals of Mathematical Statistics* 33 (3): 1187–1192. https://doi.org/10.1214/aoms/1177704481

WANG, D.–LI, V. J.–YU, H. (2020): Mass appraisal modeling of real estate in urban centers by geographically and temporally weighted regression: a case study of Beijing's core area *Land* 9 (5): 143. https://doi.org/10.3390/land9050143

WENUR, G. H.–PURHADI–SUHARSONO, A. (2020): Three-parameter bivariate gamma regression model for analyzing under-five mortality rate and maternal mortality rate *Journal of Physics: Conference Series* 1538: 012054. http://doi.org/10.1088/1742-6596/1538/1/012054

WOOLDRIDGE, J. M. (2010): *Econometric analysis of cross section and panel data* MIT Press, London.

YASIN, H.–INAYATI, S.–SETIAWAN (2022a): 3-parameter gamma regression model for analyzing human development index of Central Java Province *Barekeng: Jurnal Ilmu Matematika dan Terapan* 16 (1): 171–180. https://doi.org/10.30598/barekengvol16iss1pp171-180

YASIN, H.–PURHADI, P.–CHOIRUDDIN, A. (2022b): Estimasi parameter dan Pengujian hipotesis model geographically weighted generalized gamma regression *Jurnal Gaussian* 11 (1): 140–152. https://doi.org/10.14710/j.gauss.v11i1.33990

YASIN, H.–PURHADI, P.–CHOIRUDDIN, A. (2023): Parameter estimation and the goodness-of-fit test for the multivariate generalized gamma distribution *International Conference on Computer, Control, Informatics and its Applications (IC3INA)*: 382–387. https://doi.org/10.1109/IC3INA60834.2023.10285742

YASIN, H.–PURHADI, P.–CHOIRUDDIN, A. (2024a): Statistical inferences for multivariate generalized gamma regression model. In: BEE WAH, Y.–AL-JUMEILY OBE, D.–BERRY, M. W. (eds.): *Data science and emerging technologies. Proceedings of DaSET 2023* Springer, Singapore. https://doi.org/10.1007/978-981-97-0293-0_33

YASIN, H.–PURHADI, P.–CHOIRUDDIN, A. (2024b): Spatial clustering based on geographically weighted multivariate generalized gamma regression *MethodsX* 13: 102903. https://doi.org/10.1016/j.mex.2024.102903

ZHANG, X.–HUANG, B.–ZHU, S. (2019): Spatiotemporal influence of urban environment on taxi ridership using geographically and temporally weighted regression *ISPRS International Journal of Geo-Information* 8 (1): 23. https://doi.org/10.3390/ijgi8010023

INTERNET SOURCES

[1] BADAN PUSAT STATISTIK [BPS] (2023): *Statistics of Jawa Tengah Province* LNCS. https://jateng.bps.go.id/ (downloaded: March 2023)
[2] BKKBN (2022): Average IQ of Indonesian kids low: BKKBN *Antara News*. https://en.antaranews.com/news/265583/average-iq-of-indonesian-kids-low-bkkbn (downloaded: May 2024)
[3] KEMENDIKBUDRISTEK (2022): *Rencana Strategis (Renstra) Kementerian Pendidikan* Kebudayaan, Riset, dan Teknologi https://www.kemdikbud.go.id/main/tentang-kemdikbud/rencana-strategis-renstra (downloaded: May 2024)