Parameter estimation and hypothesis testing on geographically weighted gamma regression

Dina Eka Putri, Purhadi, and Dedy Dwi Prastyo
Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
E-mail: msdinaputri@gmail.com, {purhadi, dedy-dp}@statistika.its.ac.id

Abstract. Having applied to observations from different location, the gamma regression yields global parameters which are assumed to be valid in any location. In fact, each location may have different characteristics such that the existence of spatial effect needs to be considered. In such a case the parameters of gamma regression are less representative. Thus, Geographically Weighted Gamma Regression (GWGR) plays into role. This study aims to estimate parameters of GWGR model using Maximum Likelihood Estimation (MLE) method and numerical optimization using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Once the parameter estimation done, the hypothesis testing procedure was used to test parameter similarity between gamma regression and GWGR as well as to test the significance of independent variables within the model, both partially using Z-test and simultaneously using Maximum Likelihood Ratio Test (MLRT).

1. Introduction
A popular and broadly used method in spatial proposed by [1, 2] was so-called Geographically Weighted Regression (GWR). The GWR has become establish as a tool in investigation of spatially varying relationship but only valid for normally distributed data. In fact, not all of the data are normally distributed. Some type of response variable might follow other distributions, such as Poisson, exponential, Weibull, Gamma, and Inverse-Gaussian.

Gamma regression is a nonlinear regression used to model relationship between one or more independent variables and a positive continuous dependent variable that follows gamma distribution. Data are generally related with the geographical location where they observed, which is called spatial data [3]. Having applied to observations from different location, gamma regression yields global parameters which are assumed to be valid in any location. In fact, each location may have different characteristics such that the existence of spatial effect needs to be considered. In such a case the parameters of gamma regression are less representative.

Thus, Geographically Weighted Gamma Regression (GWGR) plays into role. GWGR is a local form of gamma regression that considers the spatial effect, i.e. longitude and latitude of point observations. Unlike conventional regression which produces a single equation to summarize global relationship among the explanatory variable and response variable, local form regression such as GWGR generates spatial data that express the spatial variation in the relationship among variables.

This study aimed to estimate parameters of GWGR model using Maximum Likelihood Estimation (MLE) method along with to test the significance of those estimated parameters. As the MLE method could not produce solution analytically, a numerical optimization method using
Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was employed to estimate the parameters of GWGR model. In addition, a Generalized Cross Validation (GCV) was used to determine the optimum bandwidth in Kernel function used as the weight function. GWGR model was expected to report better estimation than the gamma regression did. Therefore a similarity model test using $F$-test was employed to evaluate the performance of those two models. Once the parameter estimation done, the hypothesis testing procedure was used to test the significance of independent variables within the model, both partially using $Z$-test and simultaneously using Maximum Likelihood Ratio Test (MLRT).

2. Material and Methodology

2.1. Gamma Distribution

Gamma distribution is a general form of exponential distribution [4] firstly introduced by the Swiss mathematician, Leonhard Euler in the 18-th century. The gamma distribution is positive continuous distribution which is often applied in many practical areas such as meteorology, business, finance, and health. The probability density function (pdf) of gamma distribution with two parameters is given by:

$$f(y, \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-\frac{y}{\theta}},$$  \hspace{1cm} (1)

where $\alpha, \theta > 0$ and $0 < y < \infty$. Thus, $Y \sim Gamma(\alpha, \theta)$. The expected value of $Y$ according to [5] is $\mu = \alpha \theta$.

2.2. Geographically Weighted Gamma Regression (GWGR)

Gamma regression is a non-linear regression used to model relationship between one or more independent variables and a positive continuous dependent variable that follows gamma distribution. The gamma regression had been developed by [6]. Given univariate response variable $Y$ and several predictors $X_1, X_2, \ldots, X_k$, the gamma regression modeling is supposed to use log linear model so that expected value ($\mu$) is only able to be positive-valued [7]:

$$\mu_i = E(Y_i) = \exp(x_i^T \beta) = \alpha \theta_i$$ \hspace{1cm} (2)

such that $\theta_i = \frac{\exp(x_i^T \beta)}{\alpha}$. Therefore, the pdf of gamma regression is:

$$f(y_i, \alpha_i, \beta) = \frac{y_i^{\alpha_i-1}}{\Gamma(\alpha_i) \left( \frac{\exp(x_i^T \beta)}{\alpha} \right)^\alpha} \exp \left( \frac{-y_i}{\frac{\exp(x_i^T \beta)}{\alpha}} \right),$$ \hspace{1cm} (3)

with $i = 1, 2, \ldots, n$. Instead of calibrating a single equation that valid globally to all of observation, the GWGR generates separate regression for each observation using different weight for each data set. A response variable $Y$ follows gamma distribution with different parameter of $\alpha$ and $\theta$ for each location, i.e. $Y \sim Gamma(\alpha(u_i, v_i), \theta(u_i, v_i))$, where $(u_i, v_i)$ captures the coordinates location $i$ [8]. Thus, the pdf of GWGR is given by:

$$f(y_i) = \frac{1}{\Gamma(\alpha(u_i, v_i)) \left( \frac{\exp(x_i^T \beta(u_i, v_i))}{\alpha(u_i, v_i)} \right)} \times A_i$$ \hspace{1cm} (4)

in which

$$A_i = y_i^{\alpha_i-1} \exp \left( \frac{-y_i}{\frac{\exp(x_i^T \beta(u_i, v_i))}{\alpha(u_i, v_i)}} \right).$$
The assumption is that observation nearby one another have greater influence on one another’s parameter estimates than observation farther apart [9].

2.3. Spatial Effects
Spatial homogeneity caused by condition of unit in observation which is basically not homogenous. The hypothesis of Breusch-Pagan (BP) test [10] is: 

\[ H_0 : \sigma_i^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2 \]

against \( H_1 : \) there is at least one \( i \) in which \( \sigma_i^2 \neq \sigma^2 \). The BP statistics test is:

\[ BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{f} \sim \chi^2_{(k+1)}, \]

where \( \mathbf{f} = (f_1, f_2, \ldots, f_n)^T \) with \( f_i = \left( \frac{e_i^2}{\sigma^2} - 1 \right) \), \( e_i = y_i - \hat{y}_i \), \( \hat{\sigma}^2 \) is the estimated variance of \( y \), \( \mathbf{Z} \) is a standardized matrix with size \( n \times (k+1) \). The \( H_0 \) is rejected if \( BP > \chi^2_{(k)} \).

Spatial dependence indicates the effect of one location to other locations. The hypothesis being tested using Moran’s I test is \( H_0 : I = 0 \) (There is no spatial dependence) against \( H_1 : I \neq 0 \) (There is spatial dependence). The statistics test employed is:

\[ Z_I = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{Var}(\hat{I})}}, \]

where \( \hat{I} = \frac{n}{(N-1)N} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y}) \left( \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \right)^{-1}, \]

\( E(\hat{I}) = \frac{-1}{N-1} \), and \( \text{Var}(\hat{I}) = \frac{N S_1 - 3S_2 S_3}{(N-1)(N-2)(N-3)} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \right) \),

with

\[ S_1 = (N^2 - 3N + 3) \left( \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} + w_{ji})^2 \right) \]

\[ - N \sum_{j=1}^{n} \left( \sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji} \right)^2 \]

\[ + 3 \sum_{j=1}^{n} w_{ij}^2 \sum_{j=1}^{n} w_{ji}^2 \]

\[ S_2 = \frac{N^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^4}{\left( N^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right)^2} \]

\[ S_3 = \left( \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} (w_{ij} + w_{ji})^2 \right) - 2N \left( \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} (w_{ij} + w_{ji})^2 \right) + 6 \left( \sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} \right)^2, \]

(5)

2.4. Spatial Weighting
Before choosing spatial weight, distance decay and the bandwidth need to be selected. Distance calculated using Euclidean distance, while the optimal bandwidth selection might be chosen by researcher or optimized using an algorithm that seeks to maximize cross-validation score. Due to its importance, the bandwidth selection procedure has been being the researchers’ focus over
the time but there is still no method claimed to have the best performance. This study used
GCV considering the method is asymptotically optimum and do not include the parameter of
population.
Suppose $d_{il}$ is the Euclidean distance from location $(u_i, v_i)$ to $(u_l, v_l)$ and $w_{il}$ is the
geographical weight at $l$-th location for coefficient estimate at $i$-th location. Some of spatial
weighting which are frequently used for geographically weighted regression [8] are as follows:

- **Fixed Gaussian**: $w_{il}(u_i, v_i) = \exp \left[ -\frac{1}{2} \left( \frac{d_{il}}{h} \right)^2 \right]$
- **Fixed Bisquare**: $w_{il}(u_i, v_i) = \begin{cases} 1 - \left( \frac{d_{il}}{h} \right)^2, & d_{il} \leq h \\ 0, & d_{il} > h \end{cases}$
- **Adaptive Gaussian**: $w_{il}(u_i, v_i) = \exp \left[ -\frac{1}{2} \left( \frac{d_{il}}{h_i} \right)^2 \right]$
- **Adaptive Bisquare**: $w_{il}(u_i, v_i) = \begin{cases} 1 - \left( \frac{d_{il}}{h_i} \right)^2, & d_{il} \leq h_i \\ 0, & d_{il} > h_i \end{cases}$

The spatial weight employed is the kernel commonly used in nonparametric regression, see [11],
and also in statistical learning method, see [12].

2.5. Procedure
(i) Estimating parameter on GWGR model using following algorithm:
(a) Determine parameter set under $H_0(\omega)$ and under population ($\Omega$).
(b) Determine likelihood and ln-likelihood function of GWGR using MLE method.
(c) If the equation yielded is not close form, then BFGS method should be employed. The
detail of BFGS algorithm can be seen on [13].
(ii) Testing on parameter estimation both simultaneously using Maximum Likelihood Ratio
Test (MLRT) and partially using $Z$ test.

3. Results and Discussion
3.1. Parameter Estimation
The GWGR parameter estimation is conducted using MLE which aims to maximize the log
likelihood function as follows:

$$
\ln L(\alpha(u_i, v_i), \beta(u_i, v_i) | i = 1, 2, \ldots, n) = B + C + D,
$$

where

$$
B = -\ln(\Gamma(\alpha(u_i, v_i))) \sum_{i=1}^{n} w_{il} + \alpha(u_i, v_i) \ln \alpha(u_i, v_i) \sum_{i=1}^{n} w_{il}
$$

$$
C = -\alpha(u_i, v_i) \sum_{i=1}^{n} (x_i^T \beta(u_i, v_i)) w_{il} + (\alpha(u_i, v_i) - 1) \sum_{i=1}^{n} \ln y_i w_{il}
$$

$$
D = -\alpha(u_i, v_i) \sum_{i=1}^{n} y_i w_{il} \exp(- (x_i^T \beta(u_i, v_i))),
$$

Since it does not generate close form solution then numerical optimization using BFGS should
be performed. The detail of the algorithm is not described here because the limited space
available. A note necessarily to be mentioned is that the result of BFGS algorithm depends on
initial values of parameters, therefore the way how to choose the good initial point is a challenge.
An evolutionary algorithm, for example a Genetic algorithm [12], can be the alternative solution
for future research.
3.2. Hypothesis Testing

Hypothesis testing consists of three kinds of tests: (i) parameters similarity test between GWGR and gamma regression, (ii) simultaneous test of GWGR parameters, and (iii) partial test of GWGR parameters.

3.2.1. Parameter Similarity Test

GWGR method provides a feasible way for testing global gamma regression relationship for spatial data. Hypothesis being tested is the null hypothesis $H_0 : \beta_{ji}(u_i, v_i) = \beta_{j}$ and $H_0 : \alpha(u_i, v_i) = \alpha$, with $j = 0, 1, 2, \ldots, k$; $i = 1, 2, \ldots, n$, against the alternative hypothesis $H_1$: at least one of $\beta_{ji}(u_i, v_i) \neq \beta_{ji}$ and $H_1 : \alpha(u_i, v_i) \neq \alpha$. The statistics test is then constructed as:

$$F = \frac{G_1^2}{G_2^2/d_1},$$

with $G_1^2$ is the deviance of gamma regression and $G_2^2$ is the deviance of GWGR, whereas $d_1$ is the degree of freedom of gamma regression and $d_2$ is the degree of freedom of GWGR. The null hypothesis $H_0$ is rejected if $F > F(\alpha, d_1, d_2)$.

Given the parameters set under population is $\Omega = \{\beta_0, \beta_1, \ldots, \beta_k, \alpha\} = \{\beta, \alpha\}$ and parameters set under the null hypothesis $H_0(\omega)$ is $\omega = \{\beta_0, \alpha_0\}$, the deviance of gamma regression is calculated by $G_1^2 = -2\ln \Lambda = -2\ln \{L(\omega) - L(\hat{\Omega})\}$ resulted in:

$$G_1^2 = E + F + G + H,$$

with

$$E = 2 \ln \Gamma(\hat{\alpha}) - 2\hat{\alpha} \sum_{i=1}^{n} y_i^2 \hat{\beta} + 2\hat{\alpha}(n \ln \hat{\alpha})$$

$$F = 2(\hat{\alpha} - 1) \sum_{i=1}^{n} \ln y_i - 2 \sum_{i=1}^{n} \left(\frac{y_i}{\exp(y_i^T \hat{\beta})}\right)$$

$$G = 2n \ln \Gamma(\alpha_0) + 2n\hat{\alpha}_0\beta_0 - 2\hat{\alpha}_0(n \ln \alpha_0)$$

$$H = -2(\hat{\alpha}_0 - 1) \sum_{i=1}^{n} \ln y_i + 2 \sum_{i=1}^{n} \left(\frac{y_i}{\exp(y_i^T \hat{\beta}_0)}\right)$$

The $G_1^2$ follows Chi-square distribution with degree of freedom equal to $k$, i.e. the number of predictors in GWGR model.

For GWGR, the deviance is obtained by letting the parameters set under $H_0(\omega)$ is $\omega = \{\beta_0(u_i, v_i), \alpha_0(u_i, v_i), i = 1, 2, \ldots, n\}$ whereas the parameter set under population is $\Omega = \{\beta(u_i, v_i), \alpha(u_i, v_i), i = 1, 2, \ldots, n\}$.

The log likelihood under $H_0$ of GWGR model is:

$$\ln L^*(\omega) = I + J + K,$$

where

$$I = -\sum_{i=1}^{n} w_i \ln(\Gamma(\alpha_0(u_i, v_i))) + \sum_{i=1}^{n} \alpha_0(u_i, v_i) w_i \ln \alpha_0(u_i, v_i)$$

$$J = -\sum_{i=1}^{n} \alpha_0(u_i, v_i) w_i \beta_0(u_i, v_i) + \sum_{i=1}^{n} (\alpha_0(u_i, v_i) - 1) w_i \ln y_i$$

$$K = -\sum_{i=1}^{n} y_i w_i \exp(-\beta_0(u_i, v_i))\alpha_0(u_i, v_i)$$
The log likelihood under population of GWGR model becomes:

\[
lnL^*(\Omega) = L + M + N,
\]

where

\[
L = -n \sum_{i=1}^{n} w_{ii} \ln(\alpha(u_i, v_i)) + \sum_{i=1}^{n} \alpha(u_i, v_i) w_{ii} \ln \alpha(u_i, v_i)
\]

\[
M = -n \sum_{i=1}^{n} \alpha(u_i, v_i) (x_i^T \beta(u_i, v_i)) w_{ii} + \sum_{i=1}^{n} (\alpha(u_i, v_i) - 1) w_{ii} \ln y_i
\]

\[
N = -n \sum_{i=1}^{n} y_i \alpha(u_i, v_i) w_{ii} \exp(-x_i^T \beta(u_i, v_i))
\]

The parameters in log likelihood under \(H_0\) and population are then maximized using MLE by calculating first derivation and make it equal to zero. If the equation yielded is not close form, BFGS once again needs to be performed. Thus, the statistics obtained is:

\[
G_2^2 = 2(L(\hat{\Omega}) - L(\hat{\omega})) = O + P,
\]

with

\[
O = -2 \sum_{i=1}^{n} \ln(\Gamma(\hat{\alpha}(u_i, v_i))) + 2 \sum_{i=1}^{n} \hat{\alpha}(u_i, v_i) \ln \hat{\alpha}(u_i, v_i) - 2 \sum_{i=1}^{n} \hat{\alpha}(u_i, v_i) (x_i^T \hat{\beta}(u_i, v_i)) + 2 \sum_{i=1}^{n} (\hat{\alpha}(u_i, v_i) - 1) \ln y_i - 2 \sum_{i=1}^{n} y_i \hat{\alpha}(u_i, v_i) \exp(-x_i^T \hat{\beta}(u_i, v_i))
\]

and

\[
P = 2 \sum_{i=1}^{n} \ln(\Gamma(\hat{\alpha}_0(u_i, v_i))) - 2 \sum_{i=1}^{n} \hat{\alpha}_0(u_i, v_i) \ln \hat{\alpha}_0(u_i, v_i) + 2 \sum_{i=1}^{n} \hat{\alpha}_0(u_i, v_i) \hat{\beta}_0(u_i, v_i) - 2 \sum_{i=1}^{n} (\hat{\alpha}_0(u_i, v_i) - 1) \ln y_i + 2 \sum_{i=1}^{n} y_i \exp(-\hat{\beta}_0(u_i, v_i)) \hat{\alpha}_0(u_i, v_i)
\]

The \(G_2^2\) follows Chi-square distribution with the degree of freedom equal to \(tr(S)\), where \(S\) is the matrix hat described in [8].

3.2.2. Simultaneous Test The hypothesis being tested is the null \(H_0: \beta_{11}(u_i, v_i) = \ldots = \beta_{kn}(u_i, v_i) = \beta(u_i, v_i) = 0, \ j = 1, 2, \ldots, k; i = 1, 2, \ldots, n\) against the alternative \(H_1: \) at least one of \(\beta_{ji}(u_i, v_i) \neq 0\). The test performed in simultaneous test similar to previous procedure in finding the deviance of GWGR \((G_2^2)\) in parameter similarity test. The null \(H_0\) is rejected if \(G_2^2 > \chi(\alpha, df_2)\).
3.2.3. Partial Test Hypothesis being tested for partial test is $H_0 : \beta_j(u_i, v_j) = 0$ against $H_1 : \beta_j(u_i, v_j) \neq 0, j = 0, 1, 2, \ldots, k$. The statistics test for partial test of GWGR is:

$$Z = \frac{\hat{\beta}_j(u_i, v_i)}{SE(\hat{\beta}_j(u_i, v_i))}$$ (12)

If $|Z_{statistic}| > Z_{a/2}$, then the $H_0$ is rejected.

4. Conclusion

Based on this study, it is concluded that the GWGR is a local form of gamma regression model considering the geographical spatial effects, i.e. latitude and longitude coordinates. The GWGR model produces estimates that are local to each observation’s location. The parameter estimation in GWGR is performed using Maximum Likelihood Estimation (MLE) with quasi-Newton numerical optimization using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Once the parameters estimation was done, the hypothesis testing procedure was used to test the significance of independent variables within the model, both partially using $Z$-test and simultaneously using Maximum Likelihood Ratio Test (MLRT).

References

[1] Brudson C, Fotheringham A S, and Charton M 1996 Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity, *Geographical Analysis* 28(4) pp 281–298
[2] Brudson C, Fotheringham A S, and Charton M 1998 Geographically Weighted Regression: Modelling Spatial Nonstationarity *The Statistician* 47(3) pp 431–443
[3] Mei C L 2004 *Geographically Weighted Regression Technique for Spatial Data Analysis* (China: Working paper, School of Science, Xi’an Jiaotong University)
[4] Tripathi R C, Gupta C R, and Pair K P 1993 Statistical Test Involving Several Independent Gamma Distribution *Ann. Inst. Statist. Math.* 45(4) pp 773–786
[5] Casella G and Berger R L 2002 *Statistical Inference, Second Edition* (Duxbury: Pacific Grove)
[6] Al-Abood A, Bakir S, and Young D 1986 Improved Estimator for The Shape Parameter in Gamma Regression (London: Brunel University Maths Technical Papers) pp 1–27
[7] Davidian M 2005 *Applied Longitudinal Analysis* (Raleigh: North Carolina State University)
[8] Fotheringham A, Brundson C, and Charlton M 2002 *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (New York: John Wiley and Son)
[9] Mennis J 2006 Mapping the Result of Geographically Weighted Regression *The Cartographic Journal* 43 pp 171–179
[10] Anselin L 1988 *Spatial Econometrics: Methods and Models* (Dordrecht: Kluwer Academic)
[11] Rismal, Budiantara I N, and Prastyo D D 2016 Mixture model of spline truncated and kernel in multivariable nonparametric regression, *AIP Conference Proceedings* 1739 (020085)
[12] Haerdle W K, Prastyo D D, and Hafner C M 2014 Support vector machines with evolutionary model selection for default prediction, In Racine J S, Su L, and Ullah A, ed. *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometric and Statistics* (New York: Oxford University Press) pp 546–573
[13] Nocedal J and Wright S 1999 *Numerical Optimization* (New York: Springer) pp 194–201