Research Article

Applied fixed effect of Geographically Weighted Panel Regression (GWPR) with M-Estimator approach to estimate sugarcane yield data in East Java

Yani Quarta Mondiana*
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Jl Veteran, Malang 65145, East Java, Indonesia

Henny Pramoedyo
Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Jl Veteran, Malang 65145, East Java, Indonesia

Atiek Iriansy
Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Jl Veteran, Malang 65145, East Java, Indonesia

Marjono
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Jl Veteran, Malang 65145, East Java, Indonesia

*Corresponding author : Email: yqmondiana@gmail.com

Article Info
https://doi.org/10.31018/jans.v16i2.5443
Received: February 2, 2024
Revised: May 01, 2024
Accepted: May 06, 2024

INTRODUCTION

Geographically Weighted Panel Regression (GWPR) was proposed by Yu (2010) as a method utilized to analyze panel data and capture diverse relationships. It is a combination of panel regression and geographically weighted regression (GWR). The GWPR considers the variations across locations by allowing the regression coefficients to vary. This method has been widely applied in several fields, such as environmental science, agriculture, and economic science.

The GWPR uses a fixed effects panel model applied to Beijing's economic development. That study demonstrated that the GWPR model had better and more precise results than the cross-sectional GWR and panel data models (Yu, 2010). Further research on GWR panels was carried out by Cai (2014), who used the GWR panel method to determine the effect of climate variations on corn production in the United States. Danlin (2021) also uses GWR panels to model the effects of high-speed rail use in China. The study demonstrates that GWR panels is better than cross sectional GWR. Both studies supported Yu (2010) study, which concludes that GWPR had better result than cross sec-

How to Cite

Abstract
Geographically Weighted Panel Regression (GWPR), a combination of panel regression and geographically weighted regression (GWR), is used to analyze panel data and capture diverse relationship between locations. GWPR was developed on data with panel-fixed effects and applied to modeling data with spatial heterogeneity and time series. One method for estimating parameters in the GWPR is weighted least squares (WLS), which are sensitive to outliers. The present study aimed to use the M-method to estimate GWPR model parameters in data containing outliers using fixed-effect GWPR modeling for the sugar cane yield in East Java of Indonesia from 2019 to 2021. Sugarcane yield data in East Java contained outliers in several areas, including Malang, Blitar, and Ngawi Districts. Because the data contains outliers, a robust method with the M estimator was applied. The results showed that plantation areas significantly affected production in all districts. The R² of the model was 0.87, showing that GWPR model with M estimation was appropriate in predicting sugarcane yield. Based on the Akaike Information Criterion (AIC) value, the GWPR model with M estimation had better performance than GWPR model alone.

Keywords: Fixed effect model, Geographically Weighted Panel Regression (GWPR), M estimator, outlier, Weighted Least Square (WLS)
tional GWR. One method for estimating parameters in the GWPR is weighted least squares (Bruna and Yu (2016); Li and Managi (2022)), which is susceptible to outliers (Zhang and Mei, 2011). The existence of outliers causes the parameter estimation of the regression model using Least Squares method to be biased, and parameter estimation results become inefficient because of large residual values (Nugroho et al., 2020). Outliers can have a significant impact on the estimation parameters of the regression model. This bias occurs because the model tries to minimize the sum of the squared differences between the observed values and the values predicted by the model. With their extreme values, outliers can disproportionately affect this minimization process, pulling the estimated parameters towards them. The presence of outliers leads to larger residual values, which are the differences between the observed data points and the values predicted by the model. These larger residuals indicate that the model is not effectively capturing the underlying relationships in the data. Consequently, the parameter estimates become less efficient as they are influenced by these extreme observations, affecting the overall accuracy and reliability of the model.

The robust parameter estimation methods may overcome the existence of outliers. Parameter estimates in robust regression include the least absolute deviation, the S estimator, the M estimator, and the MM estimator. Robust parameter estimates have been developed in GWR modeling. This method has been applied by Harris et al. (2014) and Zhang and Mei (2011). Harris et al. (2014) used a robust, geographically weighted method to detect multivariate spatial outliers and reduce their impact on the estimates of the regression coefficients for simulated data and freshwater chemistry data for Great Britain. The research of Zhang and Mei (2011) implemented robust estimation method that used local absolute deviation (LAD) to estimate GWR parameter of simulated data. That research concluded that LAD can not offer a reliable variance estimate with parameter of simulated data.

That research revealed the modeling of robust GWR produced a better model than GWR. Erda et al. (2019) used the M estimator in GWR to model data containing outliers and produced the same conclusions as Putra (2019). According to Alma (2011), the M estimator is the simplest method and has the highest efficiency level compared to other estimators. Based on the above mentioned, this present study applied the M method to estimate parameters in sugar cane yield data containing outliers using fixed-effect GWPR modeling.

Indonesia is ranked 6th as the country with the highest sugar consumption, whose consumption reached 7.8 million metric tons in 2020. The figure for sugar consumption increased by 200 thousand metric tons in 2021, reaching 8 million metric tons. The region with Indonesia's highest sugar cane production is East Java, which comprises 38 districts/cities. An increase should also follow an increase in sugar consumption and sugar production. Sugar production in Indonesia is still insufficient for consumption, as indicated by Indonesia’s sugar import figures in 2022, which increased by 9.6% compared to 2021 (BPS, 2022).

The significant volume of sugar imports in Indonesia can serve as an example or reflection of the country’s still-weak domestic sugar sector. The weak sugar industry is driven by the low efficiency of sugar factories in Indonesia, which results in suboptimal sugar production and productivity and higher production costs. The sugar factory’s manufacturing process is highly dependent on the availability of raw materials for the sugar industry, mainly sugar cane. The sugar factory requires raw materials like sugar cane for further processing into sugar. However, the sugar cane plants required for these sugar production operations have faced restricted production in recent years. Thus, the sugar production must match the production capacity of the sugar mills available in each factory. Sugarcane production is limited due to shrinking planting and harvesting areas for sugarcane commodities. As a result, with planting and harvesting areas decreasing, it is vital to boost sugarcane production by optimizing available inputs (Harlaniingtyas and Hartatie, 2021). This study aimed to determine the elements influencing sugarcane production in East Java Province using GWPR with M estimation. The sugarcane yields were modeled through eight explanatory variables.
MATERIALS AND METHODS

This study analyzed sugar cane yield data in East Java, Indonesia, which is the primary raw material used to produce sugar. The data were collected from the Indonesia Statistic Central Agency for three years (2019 – 2021) and Statistical Of National Leading Estate Crops Commodity (2019-2021) (Directorate of General Estate Crops, 2021). The variables used in this present study consisted of eight X variables, including sugarcane plantation area, rainfall, number of farmers, sunny days, amount of fertilizer, duration of sunshine, temperature, and humidity, while Y had merely a variable (sugarcane production) (Table 1).

The procedures of data analyses were carried out as follows:

1) Describing the data and performing the Breusch-Pagan test to detect spatial heterogeneity. Given the hypotheses, the Breusch-Pagan test is expressed as follows to assess the existence of spatial heterogeneity.

   \( H_0: \sigma^2_{(u_i,v_i)} = \sigma^2_{(u_j,v_j)} = \ldots = \sigma^2_{(u_n,v_n)} = \sigma^2 \)  
   (there is no spatial heterogeneity)

   \( H_1: \text{at least one } \sigma^2_{(u_i,v_i)} = \sigma^2 , i = 1,2, \ldots, n \)  
   (there is spatial heterogeneity)

The formula for the Breusch-Pagan statistical test is as follows (Anselin, 1988)

\[
BP = \frac{1}{n} h^2 (Z'Z)^{-1}Z' \chi^2_{p+1}
\]

where

\( h_t = (\frac{e_t^2}{\hat{\sigma}^2} - 1) \)

\( Z \) : explanatory variables matrix

\( H_0 \) is accepted if \( BP \leq \chi^2_{p+1} \), while \( \chi^2_{p+1} \) is the critical value of chi-square distribution and \( p \) is the number of explanatory variables.

2) Conducting outlier detection with Z value. Z-value greater than +3 or less than -3 is considered an outlier.

3) Performing GWR model and determining longitude and altitude coordinates observation area.

4) Calculating Euclid Distance \( (d_{ij}) \) between the location to \( -i \) and location to \( -j \) with the following formula

\[
d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}
\]

5) Determining bandwidth using CV optimum criteria

\[
CV(h) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}
\]

6) Calculating weighting matrix \( (w_{ij}) \)

\[
w_{ij} = \exp \left( -\frac{1}{2} \left( \frac{d_{ij}}{h_i} \right)^2 \right)
\]

7) Performing Fixed Effect Geographically Weighted Panel Regression with M Estimation with the following algorithm

**RESULTS AND DISCUSSION**

Fig. 1 shows the sugarcane production distribution in the East Java district area. Malang Regency was the district/city with the highest sugar cane production in East Java. In contrast, four districts, Pacitan, Pamekasan, Blitar, and Surabaya Districts, had no sugar cane production from 2019 to 2021 (Fig. 1). The next step is testing spatial heterogeneity. The p-value obtained was less than alpha, indicating spatial heterogeneity characteristic (Table 2). Therefore, data can be analyzed using spatial analysis, including Geographically Weighted Panel Regression (GWPR).

Outlier detection was carried out to detect whether or not there are outliers in the data. The absence of outli-
ers in the data makes the analysis results biased. Outlier detection is carried out using the Z value. In the present study, GWPR modeling was also carried out using panel data that had been transformed (demeaning) according to the within estimator concept. Based on the detection of outliers in GWPR data, there were outliers in eight East Java districts. These districts included Malang in 2019–2021, Blitar in 2020, Batu in 2020 and 2021, Kediri in 2021, and Ngawi in 2021.

GWPR modeling in this research requires estimators that can accommodate the outlier data. This research adopted the M estimator to produce a robust model. The results demonstrated that planting area (X1) has a significant effect ($t$-value = 20.58; $p < 0.05$) on sugarcane yield while the rainfall and the number of farmers had no effect on sugarcane yield (Table 3).

The summary of the local parameter estimation model using the M-Estimator for districts/cities in East Java is presented in Table 4. The highest standard deviation was found in $\beta_7$ which indicated the tendency of temperature variation across the observation areas. Whereas $\beta_5$ (amount of fertilizer) has the smallest standard deviation, meaning the amount of fertilizer was quite even. Parameter estimates of sugarcane plantation areas had a minimum value 2.941 and maximum 7.159. Rainfall variables have a minimum value of parameter estimate -2.507 and a maximum 2.85.

Table 5 shows the significant variables that acted as predictors in most locations. Results indicated that planting areas may increase sugar cane yields. Meanwhile, the rainfall variables and the number of farmers did not affect sugarcane yields. Variable X3 was significant in three districts. Variable X6 was significant in two districts. Variable 8 was only significant in one district. When interacting with other variables, the interaction of variables X4 and X5 was significant in two districts, while the interaction between X3, X4, and X8 was significant in three districts.

The determination coefficient value was 0.87. Cheng et al. (2022) and Shaw et al. (2022) highlighted that...
Kalimantan Province showed that several factors affect the Human Development Index. Research conducted in East Java land area positively affected sugarcane productivity using multiple linear regression (Harlianingtyas and Hartatie, 2012). Another study in the East Java land area positively influenced rice production (Gamayanti, 2023). Other research shows that the GWPR model for the case of the human development index in East Java with a Fixed Gaussian weighting function was better than the global regression model (Wati and Utami, 2020). Other research shows that the GWPR model for the case of the human development index in East Java with a Fixed Gaussian weighting function was better than the global regression model (Wati and Utami, 2020). Other research shows that the GWPR model for the case of the human development index in East Java with a Fixed Gaussian weighting function was better than the global regression model (Wati and Utami, 2020). The present study showed that GWPR was successful in detecting outliers that existed in eight districts in East Java. GWPR was also suitable for analyzing variable X, which influences variable Y. The results demonstrate that planting area significantly affected sugarcane yield, while the rainfall and the number of farmers did not affect sugarcane yield. In Indonesia, the GWPR application to analyze variables influencing agricultural production has been carried out in East Java and Central Sulawesi. The other research results support the findings of Central Sulawesi research, which show that land area and harvested area, individually or interactively, influenced rice production (Gamayanti et al., 2023). Another study in the East Java land area positively affected sugarcane productivity using multiple linear regression (Harlianingtyas and Hartatie, 2012). GWPR is also used to test variables related to the Human Development Index. Research conducted in East Kalimantan Province showed that several factors affecting HDI in each of the ten regencies/municipalities. These factors are the labor force participation rate, number of health facilities, Gini ratio, population growth rate, open unemployment rate, poverty gap index and percentage of food expenditure. The coefficient of determination of the GWPR model obtains a value of 94.36% with the RMSE value of 0.1221 (Ananda et al., 2022). Other research shows that the GWPR model for the case of the human development index in East Java with a Fixed Gaussian weighting function was better than the global regression model (Wati and Utami, 2020).

The GWPR model is commonly applied to model environmental problems and transportation and innovation drivers. In the environmental field, GWPR examines the relationship between satellite-derived data, measured ground-level NO2 concentrations, and several controlling meteorological variables (Li and Managi, 2022). In transportation, GWPR has been used to analyze HSR (High-Speed Railway) station distribution data and a series of socio-economic information panel data at China’s county level. The research created four HSR accessibility indices and sought to provide insight into how access to HSR systems supports China’s county-level development (Yu et al., 2021).

Table 4. Summary of Local Parameter Estimate for GWPR model with M-Estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>1597.4</td>
<td>3174.27</td>
<td>1857.36</td>
<td>153.49</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2.941</td>
<td>7.159</td>
<td>6.57</td>
<td>0.612</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-2.507</td>
<td>2.85</td>
<td>0.69</td>
<td>0.964</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.136</td>
<td>0.92</td>
<td>0.008</td>
<td>0.223</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-94.04</td>
<td>312.243</td>
<td>36.69</td>
<td>100.69</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-6.9 x 10-4</td>
<td>7.3 x 10-4</td>
<td>3.8 x 10-4</td>
<td>3 x 10-4</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-91.38</td>
<td>-16.11</td>
<td>-54.57</td>
<td>12.49</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-679.167</td>
<td>583.346</td>
<td>235.509</td>
<td>178.325</td>
</tr>
<tr>
<td>$\beta_8$</td>
<td>-297.827</td>
<td>5.246</td>
<td>-110.654</td>
<td>47.038</td>
</tr>
</tbody>
</table>

Table 5. Significant Variables in Fixed Effect GWR with M-Estimator

<table>
<thead>
<tr>
<th>No</th>
<th>Significant Variable</th>
<th>District/city</th>
<th>Number of Significant Districts/City</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X1</td>
<td>Ponorogo District, Trenggalek District, Tulungagung District, Blitar District, Kediri District, Malang District, Lumajang District, Jember District, Banyuwangi District, Bondowoso District, Situbondo District, Probolinggo District, Pasuruan District, Sidoarjo District, Mojokerto District, Jombang District, Nganjuk District, Madiun District, Magetan District, Ngor District, Bojonegoro District, Tuban District, Lamongan District, Gresik District, Bangkalan District, Kediri City, Malang City, Madiun City, Batu City</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>X3</td>
<td>Lumajang District, Blitar City, Mojokerto District</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>X4, X5</td>
<td>Malang District, Ngaji District</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>X6</td>
<td>Tulungagung District, Nganjuk District</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>X3, X4, X8</td>
<td>Blitar district, Ngor District, Kediri District</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>X8</td>
<td>Mojokerto City</td>
<td>1</td>
</tr>
</tbody>
</table>
In innovation drivers, European research compared two local regressions, namely GWR and GWPR, to identify regions with similar innovation-driving force characteristics. The paper points to the GWPR method as a procedure to fill the gap between the GWR literature and the panel data literature. The main originality of GWPR is that it allows for studying potential spatial heterogeneity in models that control for individual heterogeneity (Musella et al., 2023).

The AIC (Akaike Information Criterion) value of the GPWR M-Estimator model in the present study was 2212. The AIC is a valuable tool in model selection, striking a balance between model fit and complexity (VanBuren et al., 2017). It allows researchers to compare different models and choose the one that best fits the data without being overly complex (Kimura, 2019). Meanwhile, in the GWPR model, the AIC was 2274.81. Therefore, in this case, the model with the AIC value of 2212.812 is preferred as it strikes a better balance between goodness of fit and model simplicity compared to the model with the higher AIC of 2274.81. Likewise, the value of the $R^2$, where the $R^2$ from the GWPR M estimator model (0.87) was greater than the GWPR (0.79). The M estimator concept involves providing a balanced weight around the average to reduce the impact of outliers on estimating results. By using a specific weighting scheme, M estimation significantly reduces the influence of outliers when updating parameter estimates, improving the estimation process’s robustness against extreme values (Wang and Lee, 2010). This weighting technique is critical in ensuring that outliers do not disproportionately influence the estimation results, making the estimation process more resistant to abnormalities in the data.

Conclusion

Sugarcane yield data in East Java in 2019-2022 contained outliers in several areas, including Malang, Blitar, and Ngawi districts. Because the data contained outliers, a robust method with the M estimator was applied. The results of the analysis showed that plantation areas significantly affected production in all locations. The $R^2$ of the model was 0.87, showing that GWPR model with M estimation is good at predicting sugarcane yield. Based on the AIC value, the GWPR model with M estimation was better than the GWPR model.

Conflict of Interest

The authors declare that they have no conflict of interest.

REFERENCES


