A multi-level spatial analysis of crime-data in Maryland

Rohit Singh

Introduction

Over the past decades, several studies have explored the trends in crime rates for different regions of the USA emphasizing the varying crime rate in the USA. This study attempts to explore the relationship between a select indicators of crime in the State of Maryland over time at the county level. Specifically, the study explores trends crime rate against property and hate crime rates. The study defines crime against property as crime committed for monetary gains such as offences of burglary, theft or arson. Such crimes have absence of force or threat of force against the victims.[1] On the other hand, hate crime is one where the intent of the offender is to cause a physical or mental harm to the victim. There have several studies that have established a linear relationship for several forms of crimes across cities of the USA (Miethe, Hughes, and McDowall 1991). In addition to this, the study also explores the relationship of number of hate crime in the counties of Maryland using Generalised linear mixed effects models. Since, the data has a panel arrangement, i.e., the data are collected by each county and year with data available for count of different forms of crimes, therefore, statistically, it makes theoretical sense to use Poisson error term instead of standard Gaussian error term for model fitting(*Econometric Analysis of Count Data* 2008).

Objectives

The study attempts to explore if the trends of crime rates (crime rate against property and hate crime rate) have linear relationship over time, across counties of the State of Maryland. For the purpose, the study, foremost, tidy the data and plot crime rates against years. Drawing insight about the trends, the study then fits linear model (OLS) along with random intercept linear mixed model and random intercept-slopes linear mixed model. The study plots the geospatial maps of the population and the change in crime rate for the counties to argue for a case of non-linear relationship between population and indicator of crime. Therefore, the study explores if the reported incidents of hate crimes and property crime are changing over time for different counties in the State of Maryland. For the estimation, the study deploys poisson count model under Generalised linear model and Generalised linear mixed effects model. Finally, the study presents the findings of the GLMER modelling process using illustrative tools.

[1] The study borrows the definition of various nature of crime from the Unforim Crime Reporting Handbook, published by U.S. Department of Justice.(https://ucr.fbi.gov/additional-ucr-publications/ucr_handbook.pdf)

Data

The data for the study have been sourced from the Maryland Statistical Analysis Center, an Open Data Portal by the State Government of Maryland, USA[2]. The Center maintains a record of the several forms of crimes, along with other demographic variables at the county level, from 1975 till 2020. The study uses the county-level nested panel data-set for the regression analysis.

Given the nested nature of the data-set, the study uses hierarchical modelling to ascertain if the indicators of crime (hate crime rate and crime rate against property) follow a linear trend in the State of Maryland from 1975 to 2020, the period of the study. For the analysis, foremost, the study selects the relevant indicators and counts of crime i.e., the crime rate against property, rate of hate crime, the count of reported crime against property and the count of reported hate crime across all the counties in the Maryland. Along with these variables, the data on demographic variable-population for all the counties in the Maryland is collated.

Before, the study fit in a linear mixed effects model, it re-scales the Year variable so that the models don't fail to converge. The LMER model fits well when the intercept term is zero. LMER model is particularly used to ascertain if a linear trend through time can predict crime rates[3]. Using the LMER model, the study estimates the a slope coefficient for the entire state as well as a coefficient for each county). Similar logic is used to fit in the GLMER model on the count data for reported crimes.

Results

The study plots the linear fit graph for each county and examines how hate crime rate and crime rate against property change over time along with the corresponding linear fit for each county in the State of Maryland over the period of study. The study observes that almost all counties have different intercept and slope in the plots of hate crime rate and crime rate against property, as shown in Figure 1. While, the study observes mixed trends in linear fits for hate crime rate, the trends in linear fits for property crime are downward sloping for most of the counties. The varying trend across the period of study stand and differences in the crime rate at the study of the period of study, establish a basis for exploring linear effect mixed effects model with both random intercept and random slopes for data fitting.

The counties in the State of Maryland have different population sizes, therefore there is a theoretical justification for incorporating population as one of the control variables in the specification. In the linear model and linear mixed effects model, the study estimates the following specifications (matrix notations):

$$Crimerate = X\beta + \epsilon - (1)$$

$$Crimerate = X\beta + Zu + \epsilon - (2)$$

In the specifications above, Crime rate is the dependent variable, which is captures crime rate against property and hate crime. β is the unknown vector of fixed effects, u is the unknown vector of random effects and ϵ is the unknown vector of Gaussian random errors in specification (2).
Estimating specification (1) using OLS, renders estimates as shown in Table 1. The columns 1 and 4 of the Table 1 show that there is a significant decreasing trend of both property crime rate and hate

[2]The data may be downloaded from the Open Data Portal by the State Government of Maryland. https://opendata.maryland.gov/Public-Safety/Violent-Crime-Property-Crime-by-County-1975-to-Pre/jwfa-fdxs

[3]https://github.com/jroy042/MixedEffectsAreSometimesTerrible

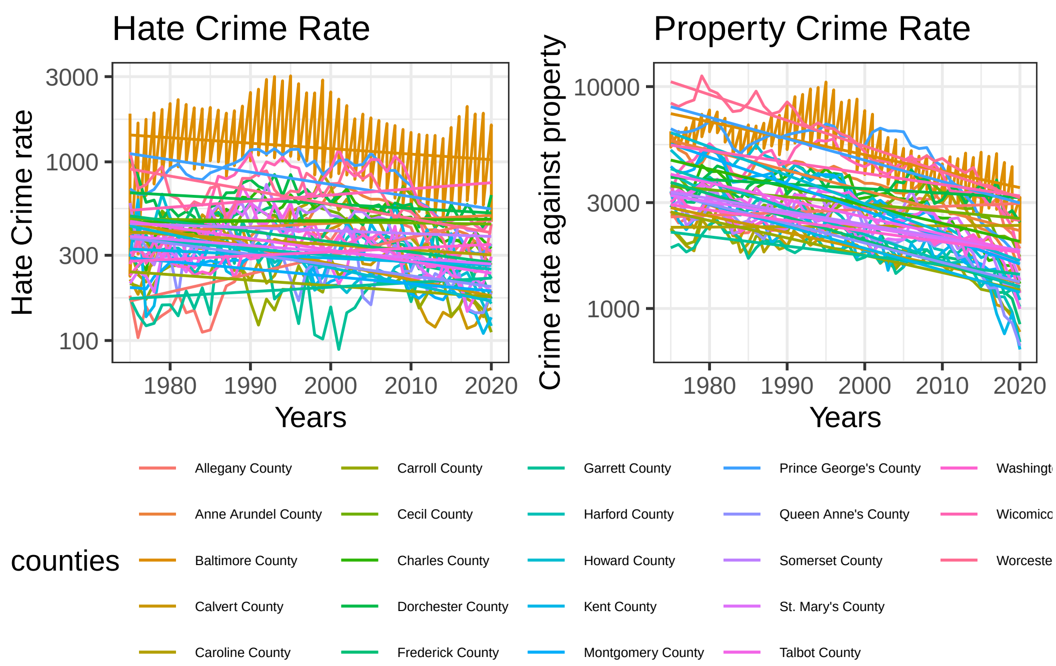

Figure 1: Relationship between crime-rates and time for counties of Maryland

crime rate in the counties over time. However, this is a pooled regression and given nested nature of the data, the result would only be robust if the same is confirmed when the study deploys LMER models on the data. Indeed, so turns out to be the case, the LMER model (with random intercept) fit, in the columns 2 and 5 show that there is a statistical negative relationship of time and the crime variables. Lastly, the LMER model (with random intercept and slope)[4] fit, in the column 3 and 6, also confirm a statistical negative relationship between the regressor and regressand.

Next, the study performs analysis of variance for the LMER models. The results of the the analysis of variance are presented in the Table 2 and Table 3 for property and hate crime rate respectively. The Chi-square value is statistically significant, favoring mixed effects model with random intercept and slopes model over only random intercept model. This is also confirmed by the lower log-likelihood value, lower AIC and BIC. This implies LMER with random intercept and slope model does better job in explaining the variability present in the data.

In the Figure 2, the study plots the spatial map of the State of Maryland with units as its counties. The figure consists of a spatial plot of total population of the each county in the Maryland and slope coefficient of the variable Year from the LMER with random intercepts and random slopes. The aim of the plot is to visualize if there are non-linear patterns of clusters of crimes vis-a-vis the population. A close comparison of the population's spatial distribution, reveals that population might have an effect on the crime rate across the counties. The Figure 2 suggest a non-linear relationship between population and crime in some of the counties. Therefore, the study use a non-linear model to estimate the specifications (1) and (2).
Statistical theory suggest for use of parallel of LM i.e., GLM with poison error term for the model

[4]It can be observed that the crime rate varies by county. The study incorporates this observation as the heuristic reason for inducing Year as both as random and fixed effect in the model.

Table 1: Linear Model and Linear Mixed Effects Model estimates

	1	2	3	4	5	6
(Intercept)	8426.87***	5244.27**	−3567.44	471.70	−25.65	−295.64
	(1976.64)	(1610.88)	(2439.07)	(615.77)	(402.89)	(473.57)
Year	−42.77***	−46.81***	−55.76***	−2.60**	−3.32***	−3.48***
	(2.88)	(2.52)	(8.25)	(0.90)	(0.71)	(0.98)
log(Population)	−415.35*	−94.81	691.28**	−9.20	47.73	71.58+
	(180.16)	(142.12)	(216.21)	(56.12)	(35.71)	(42.06)
Num.Obs.	1104	1104	1104	1104	1104	1104
R2	0.764			0.635		
R2 Adj.	0.759			0.626		
Log.Lik.	−8904.886			−7617.309		
F	145.909			78.073		
RMSE	770.53	771.72	618.49	240.04	240.19	235.37

1: LM for property crime rate; 2: LMER model with intercept for property crime rate
3: LMER model with intercept and slopes for property crime rate; 4: LM for hate crime rate
5: LMER model with intercept for hate crime rate
6: LMER model with intercept and slopes for hate crime rate
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

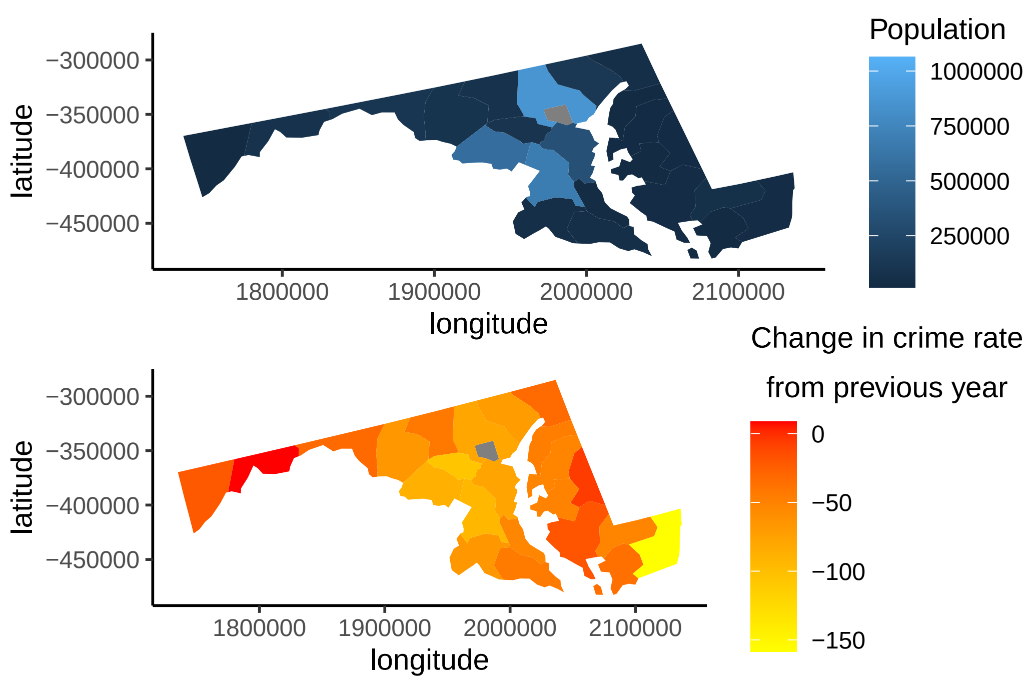

Figure 2: Spatial maps comparison

fitting. For the purpose, the dependent variable that the study considers is the reported count of incidents of crime data, i.e., reported counts of incidents of hate crime and property crime.
Further, in the study uses poisson count data models to measure relationship between counts of reported incidents property crime and hate crime with time using Generalised Linear Model and Generalised Mixed-effects Model (GLMER). Since, the data qualifies to be classified as *'repeated measure'*, therefore, fitting GLMER has the theoretical justification. The study, fits GLM and GLMER (intercept and slope model) to the specifications (1) and (2) and the results are presented in Table 4. Examining the estimates, the study observes that GLM estimates for Year are negative and statistical significant on both count of reported hate crime and count of reported property crime. However, this is not the only direction of relationship the study puts forth. Examining the GLMER estimates, the study notes that there is a statistical significant and positive coefficient of Year for count of reported hate crime, while it is insignificant for the count of reported property crime.

Table 2: ANOVA (Property Crime Rate-LM and LMER model comparison)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMER (Intercept)	5	17955.2	17980.2	-8972.6	17945.2			
LMER (Intercept-Slope)	6	17589.8	17619.8	-8788.9	17577.8	367.4	1	0

Table 3: ANOVA (Hate Crime Rate-LM and LMER model comparison)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
LMER (Intercept)	5	15357.4	15382.4	-7673.7	15347.4			
LMER (Intercept-Slope)	6	15350.3	15380.4	-7669.2	15338.3	9	1	0

The study find contrasting directions of coefficients of Year for count of reported hate crime, and therefore analysing the variance of the two models becomes paramount for statistical conclusions. The Anova results for are presented in the Table 5. The results clearly indicate statistical choice in favor of GLMER model. Additionally, the value of log-likelihood, AIC and BIC also indicate that GLMER model captures the variability in the data better than GLM.
The study used GLMER ransom intercept and slopes model to fit the reported count of hate crime data. For this to be visually true, there must be some variability in the slope coefficients of the variable Year for each county in the State of Maryland. A visual plot of the variability in the slopes is plotted in the Figure 3. In the plot, the vertical axis records the county while the horizontal axis records the value of Year variable coefficient from the GLMER model fit of the reported count of hate crime data. The variability strongly support the study's decision of fitting a random slopes and random intercept model.

Table 4: Generalised Linear Mixed Effects Model estimates

	GLM (hate)	GLM (Property)	GLMER (hate)	GLMER (Property)
(Intercept)	5.52***	7.88***	5.80***	7.96***
	(0.01)	(0.003)	(0.28)	(0.29)
Year	−0.005***	−0.01***	0.006***	−0.002
	(0.000 06)	(0.000 03)	(0.001)	(0.002)
Population	0.000 000 5***	0.000 000 4***		
	(1×10^{-8})	(4×10^{-9})		
Num.Obs.	1104	1104	1104	1104
Log.Lik.	−155 923.646	−282 010.244		
F	102 534.317	456 756.212		
RMSE	1523.20	3984.54	1511.56	3987.94

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Anova (Hate Crime Rate-GLM and GLMER model comparison)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
GLMER	5	291571.3	291596.4	-145780.7	291561.3			
GLM	25	311897.3	312022.5	-155923.6	311847.3	0	20	1

Limitations

There are some limitations of the study that should be noted at this point of time. The focus of this study was to explore linear and non-linear relationship between the forms of crimes over time and restricting to only these two possible form of relationships is itself a limitation. This is because the study does not explore spatial-dependence of crime across counties. All counties share their boundaries with other counties and counties at the boundary of the states share thier boundaries with counties from other states. This means there is an internal spatial dependence along with an external spatial dependence, the study should correct coefficients for (based on the fact, how it defines boundary criteria). For instance, crime rate could be higher in counties towards the periphery of the state because there is a spill over of criminal activity (cause of crime) from other states. Further, some counties of Maryland share their boundaries with the Atlantic ocean, the nature of crimes could be different in these counties.

Secondly, the study did not include several plausible control variables, those may create a co-founding effect leading to omitted variable bias. These could be factors like aspects of culture, ethnicity, religion adherence rate, sex ratio, policing, social support, unemployment and other demographic and socio-economic variables. The study did include population, but surely, non-inclusion of others may cause issues in perfectly measuring causal relationship.

For reasons of heterogeneous population and socio- economic differences across counties and states, the job of policy marker would be difficult with only such naive analysis of crime data.

Lastly, for the measurement, the study did not consider several non-linear estimation techniques like GAMS or Bayesian modelling. These methods may allow the study to estimate credible confidence

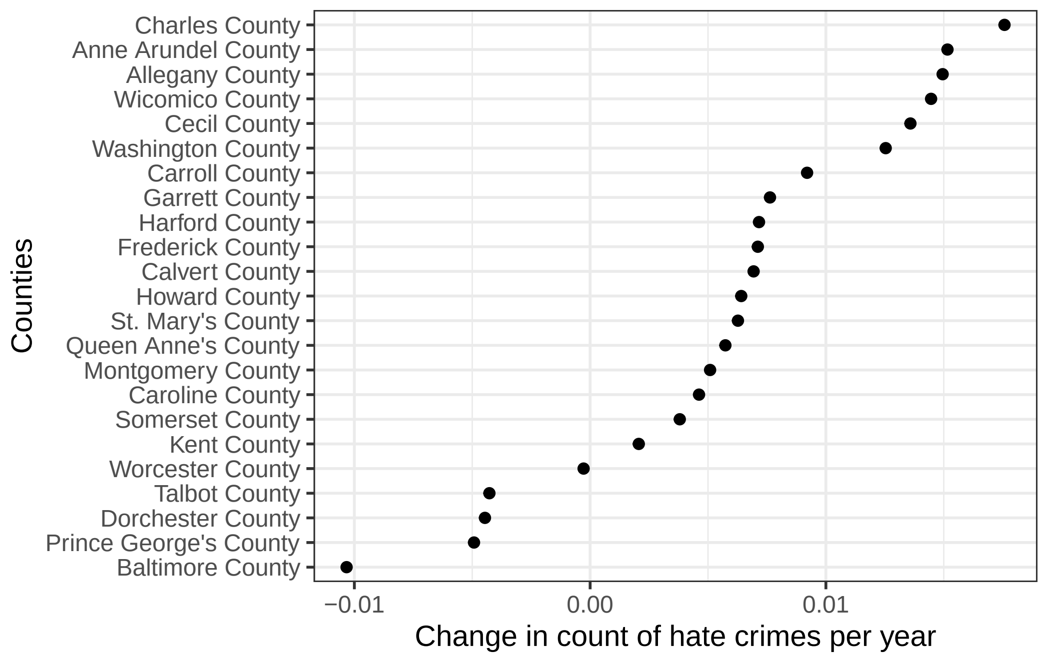

Figure 3: Change in Reported Hate crime per year

interval for random effects.

Conclusion

The aim of this study was to examine the relationship of hate crime and crime against property with time in counties of the State of Maryland. For the purpose, the study used the crime data maintained by the Maryland Statistical Analysis center. The study presented a graphical analysis of the trend of hate crime rate and property crime rate over time for all counties. The graphical analysis indicated that trend varies with county and has different intercept.

The study fitted linear model onto the crime date to ascertain the relationship. The variable year did have a strong linear relation with the crime rate in linear fit, LMER with random intercept and LMER with random intercept and slope model. However, the population variable did not conform to the linear structure of the fit.

The study then presented a spatial graphical analysis of the population and crime rate to investigate if there is an non-linear relationship that can fit the data. For the purpose, the study did fit GLM and GLMER with random intercepts and random slopes to the data and analysis of variance did confirm a rise in hate crime in the counties of Maryland. This result stood in contrast to the previous findings of the data fit.

Lastly, the study presented the graphical representation of change in count of hate crimes per year obtained from the GLMER to empahise the variability of hate crime rate across the counties.

Appendix- A

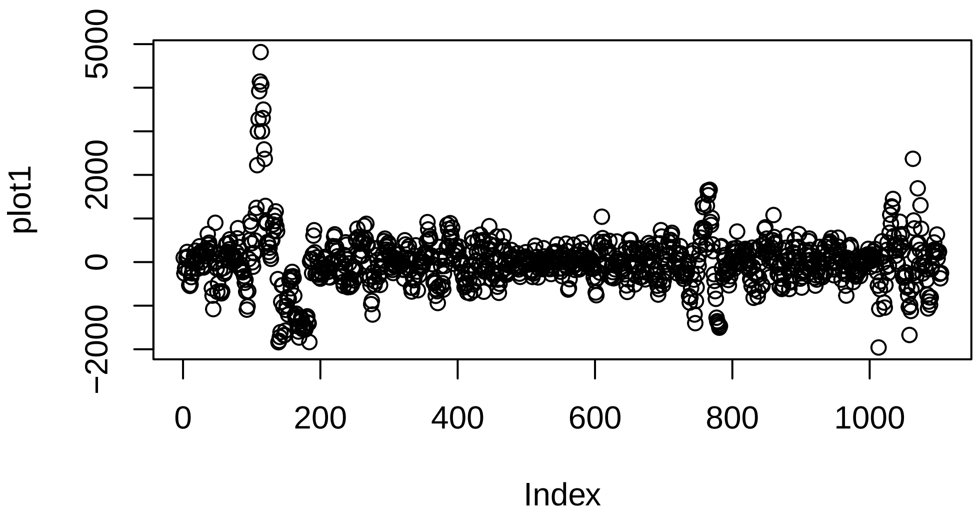

Figure 4: Residual plots of LMER and GLMER

Figure 4 plots the residual plots of the LMER and GLMER models. It is clearly evident that residuals are symmetrically distributed around zero.

- Plot-1 is the residual plot of LMER-property crime rate.
- Plot-2 is the residual plot of LMER-hate crime rate.
- Plot-3 is the residual plot of GLMER-hate crime rate.

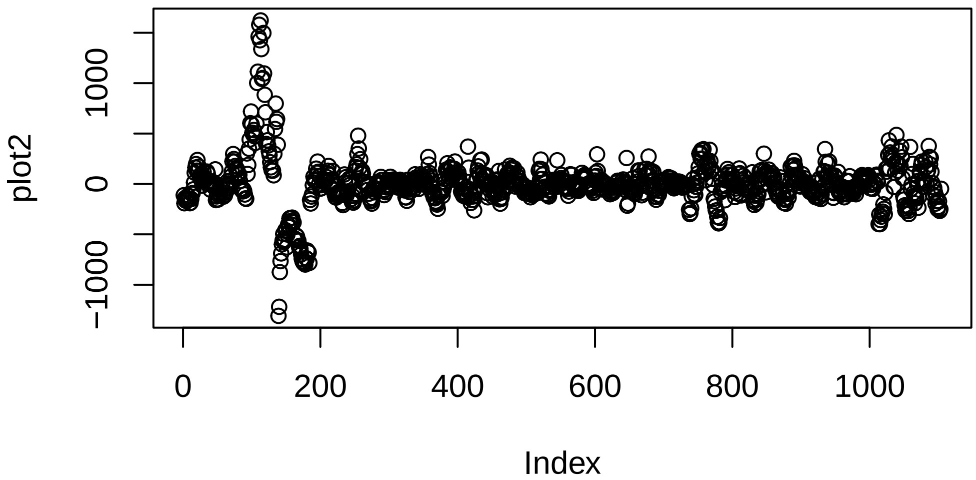

Figure 5: Residual plots of LMER and GLMER

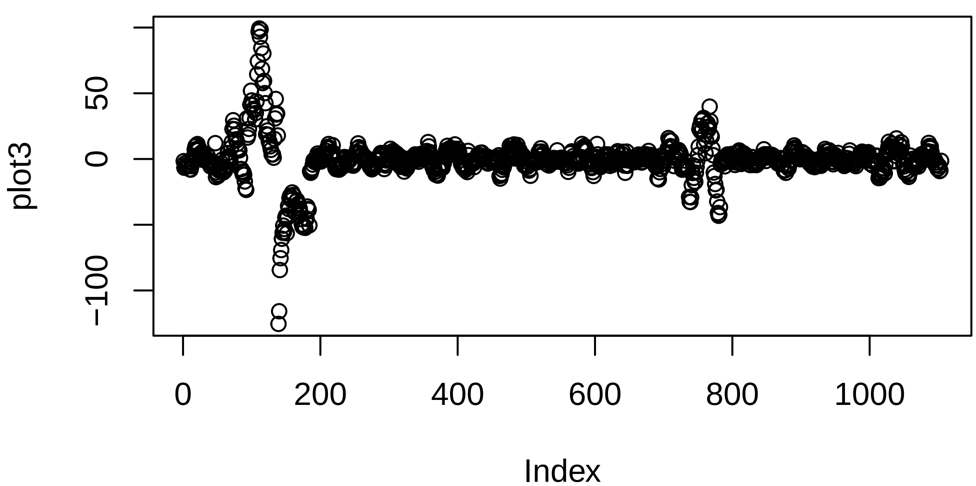

Figure 6: Residual plots of LMER and GLMER

References

Econometric Analysis of Count Data. 2008. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78389-3.

Miethe, Terance D., Michael Hughes, and David McDowall. 1991. "Social Change and Crime Rates: An Evaluation of Alternative Theoretical Approaches." *Social Forces* 70 (1): 165. https://doi.org/10.2307/2580067.