12-2020

Quantifying the Simultaneous Effect of Socio-Economic Predictors and Build Environment on Spatial Crime Trends

Alfieri Daniel Ek
University of Arkansas, Fayetteville

Quantifying the Simultaneous Effect of Socio-Economic Predictors and Build Environment on Spatial Crime Trends

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Science in Statistics and Analytics

by

Alfieri Daniel Ek

University of Belize
Bachelor of Science in Mathematics, 2018

December 2020
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Jyotishka Datta, PhD
Thesis Director

Grant Drawve, PhD
Committee Member

Samantha Robinson, PhD
Committee Member

Giovanni Petris, PhD
Committee Member

Abstract

Proper allocation of law enforcement agencies falls under the umbrella of risk terrain modeling (Caplan et al., 2011, 2015; Drawve, 2016) that primarily focuses on crime prediction and prevention by spatially aggregating response and predictor variables of interest. Although mental health incidents demand resource allocation from law enforcement agencies and the city, relatively less emphasis has been placed on building spatial models for mental health incidents events. Analyzing spatial mental health events in Little Rock, AR over 2015 to 2018, we found evidence of spatial heterogeneity via Moran's I statistic. A spatial modeling framework is then built using generalized linear models, spatial regression models and a tree based method, in particular, Poisson regression, spatial Durbin error model, Manski model and Random Forest. The insights obtained from these different models are presented here along with their relative predictive performances. These inferential tools have the potential to aid both law enforcement agencies and the city in properly allocating resources required for such events.

Acknowledgements

First and foremost, I would like to give thanks to the Lord almighty for granting me this lifetime accomplishment.

I would like to thank everyone at University of Arkansas for the enormous support provided to me throughout the two years of my masters studies, especially the department of Mathematical Science. My thanks to every faculty member of the department of Mathematical Science whom I had the pleasure to work with, take a class with or simply corresponded with. A special thanks goes to my thesis adviser/instructor and friend Dr. Jyotishka Datta for guiding me through the process of writing this thesis and wonderful lessons instilled onto me, I truly cherish them. My fullest gratitude towards my family back in Belize for the unconditional support and motivation, they are my pillar and source of perseverance. My final thanks goes to both my beloved brothers Luis Javier Ek Jr and Luigi Aldair Ek, whom from a young age and until present continue to challenge and guide me, thanks for the support you both have given me these past years.

Table of Contents

Chapter 1

Introduction

Over the last two decades, law enforcement agencies are relying more and more on statistical tools to build an objective criminal justice system, leading to a meteoric rise of "predictive policing", loosely defined as "*the application of analytical techniques - particularly quantitative techniques - to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions*" (Perry et al., 2013). The proposed algorithms and methods attempt to uncover and exploit different aspects of crime activities data. For example, Gotway & Stroup (1997) use a spatial generalized linear model, that has been extended both by considering the temporal pattern as well as a non-linear modeling approach using generalized additive modeling in ST-GAM or LST-GAM (Wang & Brown, 2012). In a series of papers, Mohler et al. (2011, 2013, 2015) propose a self-exciting point process model that treats near-repeat nature of crimes (Townsley et al., 2000) as aftershocks of an earthquake. This is the main driving force behind the popular crime forecasting software called PredPol (`https://predpol.com/`) that has been since adopted by many policing agencies over the US.

Apart from increasing the accuracy of prediction of future crime, it is also important to understand which geographical factors significantly contribute to crime that can inform a plan for allocating resources or making policy changes to either counteract the effect of 'risky' place or increase the intensity or presence of a 'protective' place. This is also closely related to the goal of ensuring that a prediction rule that does not suffer from algorithmic or systemic biases. This is particularly important, as with the increase in complexity and use of such data-based tools, there is an increasing concern of reducing the racial disparities in predictive policing, while producing dynamic and real-time forecasts and insights about spatio-temporal crime activities. For example, using a combination of

demographically representative synthetic data and survey data on drug use, Lum & Isaac (2016) point out that predictive policing estimates based on biased policing records often accentuate the racial bias instead of removing it. A natural solution seems to be the risk terrain modeling (RTM) framework of Caplan et al. (2011), that uses a simple but interpretable approach. In RTM, a separate map layer is created for each predictor, that are then combined to produce a composite map where contribution or importance of each factor can be evaluated in a model-based way.

We start with a brief review of the existing statistical methodology behind the most common crime forecasting tools.

1.1 Literature Review

Self-exciting Point Process: One of the popular statistical approaches to modeling criminal activities is self-exciting processes (Mohler et al., 2011, 2013, 2015) that is characterized by the increasing probability of repeated events following an event, similar to aftershocks of an earthquake. Here the intensity of a discrete-time point process (criminal activities, in this context) is determined as a log-Gaussian Cox process (LGCP) whose intensity is self-excited by occurrence of many events in a short time-window.

Generalized Additive Modeling for Spatio-temporal Data: Wang & Brown (2012) developed a more sophisticated model using a generalized additive modeling for spatio-temporal data (ST-GAM) that can be thought of as an extension of grid-based regression approaches that can account for non-linear relationships. Here, spatio-temporal features include previous crime activities, socio-economic and built-environment features at the grid-cell resolution indexed over time, and Wang & Brown (2012) showed that their method outperforms spatial Generalized Linear Model (GLM) (Gotway & Stroup, 1997) where temporal information is not incorporated.

Risk Terrain Modeling: Risk terrain modeling, henceforth abbreviated as RTM, (Caplan

et al., 2011, 2015; Drawve, 2016) is a class of statistical methods that combines geographic features such as built-environments and socioeconomic variables in a supervised learning set-up to provide insights and forecasts for crime activities at a chosen grid-level based on the proximity to features and social factors. A typical RTM approach involves three steps: (1) identifying potentially relevant factors for the spatial varying response variable, (2) assign a value for each factor considered for each location or grid-cell spanning a common geography, and (3) combine the factor-specific raster maps in a supervised regression framework so that each factor can be judged in terms of its relevance for the crime outcome. There are three key advantages of risk terrain modeling approach over the LST-GAM or Hawkes process based algorithms. Firstly, the underlying statistical methodology for RTM immediately provides interpretability to the factors influencing spatial clustering of crime or other response variables. Secondly, it alleviates some of the racial disparity concerns by moving the focus of the modeling approach from people to places. Finally, the raster-map based modeling framework lets us easily incorporate different machine learning and statistical tools of choice depending on their performance for a given jurisdiction. In this thesis, we use Poisson GLM, spatial error model and random forest, but it is straightforward to add any number of methodologies to the mix and choose the best performing method or combine the disparate tools in an ensemble learning framework.

While these developments have been mostly focused on crime prediction and prevention, there is relatively less emphasis on other spatial events such as mental health calls that also require resource allocation from the law enforcement agency or the city. The goal of this thesis is to extend the powerful and interpretable statistics and machine learning methodologies under the general umbrella of risk terrain modeling to the geo-spatial predictive modeling of mental health call locations in Little Rock, AR.

The outline of the thesis is as follows: in Chapter 2, we describe the modeling

approach and the different methodologies used in developing the risk terrain for mental health calls. Next Chapter 3 illustrates the spatial clustering and other descriptive features of the data and demonstrates the performances of the proposed framework. Finally, in Chapter 4, we provide some new directions for research in this area.

Chapter 2

Spatial Forecasting

2.1 Modeling Approach

Our spatial modeling and forecasting framework is similar to RTM, with a key difference being the underlying statistical methodologies. In this thesis, we use the following methodologies and compare both the important predictors chosen by the model as well as their predictive performance for forecasting mental health incidents in Little Rock, AR:

Poisson Generalized Linear Model The Poisson regression model belongs to a family of regression models called the generalized linear model (GLM). As a special case of the GLM family, the fitted Poisson regression model uses $\eta_i = \ln(\lambda)$ as canonical link and is of the form:

$$\hat{y_i} = g^{-1}(x_i^T \hat{\beta}) = \mathrm{e}^{x_i^T \hat{\beta}}.$$

Among several link functions commonly used with the Poisson distribution, the log link function ensures that $\lambda_i \geq 0$ which is crucial for the expected value of a count outcome of response variable (mental health incidents) (Montgomery et al., 2006). In terms of model interpretation, parameters may be interpreted in a probabilistic sense which arises as an advantage from the fact that Poisson regression belongs to the GLM family. This suggest that significant factors present in the fitted model may be explained in strict probabilistic terms with respective levels of uncertainty.

Random Forest Random forest (Breiman, 2001) falls into the non-linear/non-parametric category of supervised learning approaches known as decision trees. Decision trees are particularly known due to their inherent ease of use and interpretability in both cases of regression and classification problems. For regression problems, decision trees divide the predictor space into J distinct and non-overlapping regions,

$R_1, R_2, ..., R_J$ also known as terminal nodes or leaves using the training data through a recursive binary splitting procedure. Note that a threshold is implemented onto the recursive binary splitting procedure at each step to ensure that the process ends when the number of observations at a given split is less than the threshold. In addition to the preceding criteria, the aim is to obtain terminal nodes that minimize the residual sum of squares:

$$\sum_{j=1}^{J} \sum_{I \in R_j} (y_i - \hat{y}_{R_j})^2.$$

The results obtained are likely to over-fit the data due to the complexity of the resulting tree so, a cost-complexity pruning procedure is implemented to find a sub tree which minimizes the objective function:

$$\sum_{j=1}^{|T|} \sum_{i:x_i \in R_j} (y_i - \hat{y}_{R_j})^2 + \alpha |T|,$$

thereby reducing the variance at the cost of little bias for better interpretation. As a preventative measure to not over-fit the training data and control the length of the tree, the penalty factor α is added to $|T|$(the number of terminal nodes). Any observation that fall into the R_i^{th} region is simply the mean response of the R variables from the training data set.

Single Decision trees however are not as competitive when compared to other forms of linear or non-linear supervise learning models. One solution to build a more robust decision tree is known as Random forests. Random forests builds B number of trees to improve its performance using bootstrapped samples from the training data in a strategic manner that decorrelates the trees and the final prediction is done by averaging the prediction from each individual trees. In the process of building

each decision tree, at every stage or split, a random sample of size $m = \sqrt{p}$ predictors are chosen as candidates from the pool of p predictors. As a result, strong predictors do not influence the building order of every tree (making them not look alike). This process decorrelates the trees, as on average $\frac{p-m}{p}$ of the splits would not have such strong predictors thus reducing the variance and improving results. We refer the reader to James et al. (2014) for an in-depth discussion of random forest.

Spatial Econometric Model: Spatial Durbin Model Data containing a location/geographic component contain spatial dependencies among observations which may lead to spatial relationship. Spatial relationships occur not only in the dependent variables (response variable), but also independent variables (covariates) and residuals terms (ϵ). The proper terms defining spatial relationships among dependent variables, independent variables and residual terms are known as endogenous interaction, exogenous interaction and error interaction respectively .A model that accounts for all spatial relationship is the **Manski model** [1], with the form:

$$\mathbf{Y} = \delta\mathbf{W}\mathbf{Y} + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \mathbf{u}; \quad \mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \epsilon. \tag{2.1.1}$$

Here δ is known as the spatial autoregressive coefficient, λ is the spatial autocorrelation coefficient, $\mathbf{W}$ represents the spatial weighted matrix that describes the spatial configuration of the unit samples, $\mathbf{X}$ is a matrix of exogenous variables or covariates and lastly θ and β are unknown parameters to be estimated that explain the contribution of each predictor and their spatially lagged version(Elhorst, 2014).

For the purpose of this thesis, both Manski and spatial Durbin error models were fitted onto the mental health spatial data. The Manksi model otherwise known as the General nesting spatial model in Fig. 2.1 models the spatial events (mental health incidents) as a function of endogenous interactions (neighboring values or spatial

[1]The Manski model is also known as the Generalized Nesting Spatial Model(GNS) (Elhorst, 2014)

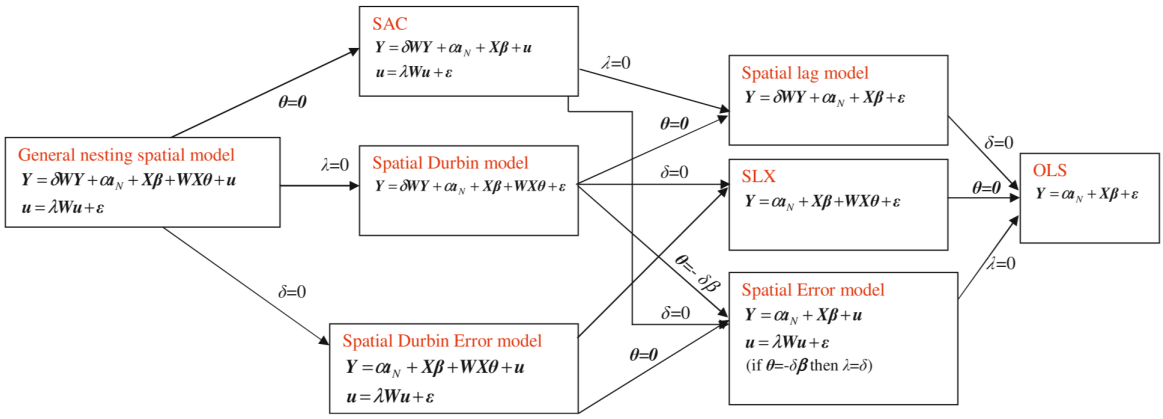

Figure 2.1: Taxonomy of Spatial models, reproduced from (Halleck Vega & Elhorst, 2015).

lags), exogenous interactions(build environment, social factors etc.) and error interactions (spatial autocorrelation& spatial heterogeneity). The spatial Durbin Error Model is a special case of a Manski model with $\delta = 0$, thus having the endogenous interactions removed. Spatial Durbin Error Model is of the form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{WX}\theta + \mathbf{u}; \quad \mathbf{u} = \lambda \mathbf{Wu} + \epsilon. \tag{2.1.2}$$

Chapter 3

Analyzing mental health incidents in Little Rock

3.1 Descriptive Statistics

3.1.1 Evidence of Clustering: Moran's I

The underlying assumption at the start of this study was that mental health incident in Little Rock were rather present as concentrated groups (*i.e.* Clusters) rather than occurring at random. To put matters into visual perspective, see Fig. 3.1 where panel 1 represents the actual geographic position of recorded 2018 mental health incidents in Little Rock and panel 2 represents the same number of incidents but simulated as if they were of random occurrence following an uniform spatial distribution. Following Fig. 3.1, it can clearly be seen the presence of concentrated zones of mental health incidents when comparing both panels. Such remarks may be interpreted as being subjective, therefore rather than relying on visual senses to identify clustered and non-clustered regions; a measure of spatial auto correlation was introduced to test the initial assumptions. In proper statistical terminology, the null hypothesis follows that mental health incidents are randomly distributed across the area of study (Little Rock) and the alternative hypothesis was that mental health incidents were more clustered than as expected from usual randomness.

It must be noted that the clustering terminology refers to the whole spatial pattern described by a global statistic for spatial auto - correlation. In order to properly identify the clustered and non-clustered regions, a specific application for a LISA (Local Indicator of Spatial Association) must be implemented. LISA is any statistic that provides the extent of significant spatial clustering of similar values around a given observation (*i.e* Local Spatial Statistic). It also establishes the connection between the local and global statistic for spatial association having the sum of all local spatial statistics be proportional to the

global statistic thereby allowing the decomposition of global indicators. (Anselin, 1995)

Among a handful number of global tests for spatial auto correlation including Geary's C and the global Getis-G, Moran's I is perhaps the most common global test, and is implemented in almost all common spatial toolboxes for testing auto-correlation (Bivand et al., 2008). Spatial auto-correlation quantifies the degree to which similar features cluster and identifies their location. In the presence of spatial auto-correlation, we can predict the values of observation i from the values observed at $j \in N_i$, the set of its proximate neighbors (Pebesma & Bivand, 2019). As in typical correlation, Moran's I value generally ranges from -1 to $+1$ inclusively as a result of having a normalizing factor, $\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$ (Boots, 2001). The contrast between spatial auto correlation Moran's I and Pearson/Spearman's correlation lie in the presence the spatial weight matrix in Moran's I statistic. The inclusion for the spatial weight matrix in Moran's I enables the possibility of obtaining extreme values greater than the usual $(-1, 1)$ bounds depending on the structure/composition of the weight matrix. Extreme Moran values are obtained via the relation between the min and max eigen values from the spatial weight matrix, for a thorough discussion regarding extreme values we refer the reader to (de Jong et al., 1984) .

A negative and significant Moran's I value represent negative spatial auto-correlation indicating dissimilar values are next to each other. A positive and significant Moran's I value represent positive spatial auto-correlation indicating evidence of clustering of liked values.

In order to apply the spatial auto correlation test (both Global and Local Moran's I) onto the Spatial Data, two critical prerequisites steps had to be executed. Steps included the identification of the k nearest neighbors then assigning their respective weights using the package **spdep** (Bivand & Wong, 2018). To identify both prerequisites, a fishnet of grid cell size of 1000m by 1000m representation from Little Rock containing all the necessary attributes for the analysis previously created was used after undergoing a centroid

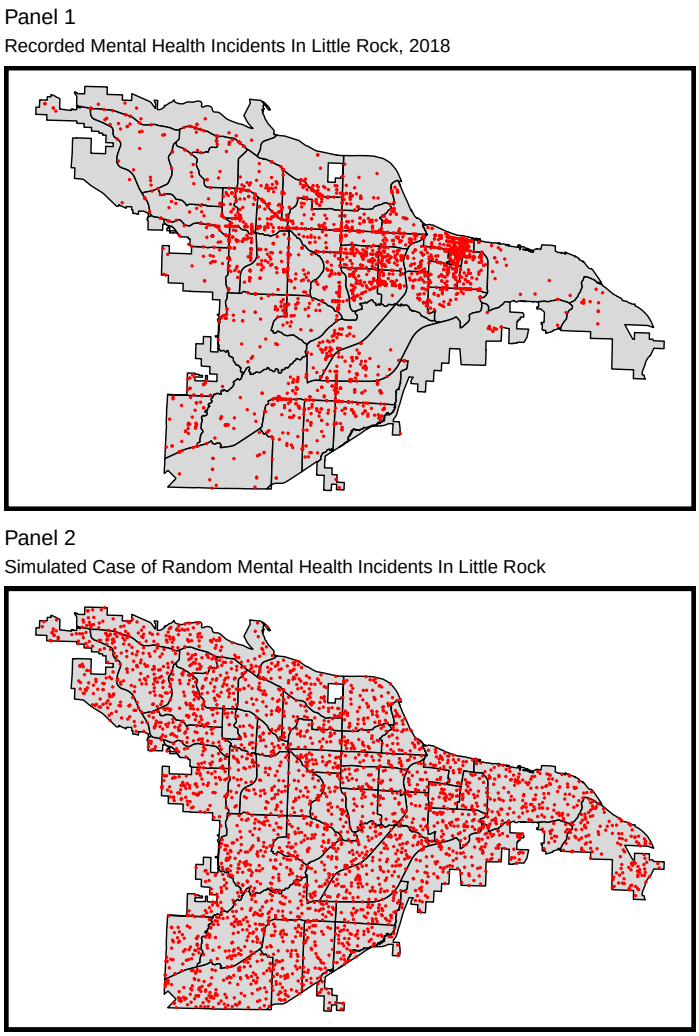

Figure 3.1: Panel 1 showing observed mental health incidents in Little Rock in 2018. Panel 2 shows distribution of simulated mental health incidents following a Uniform distribution, keeping the total number of incidents fixed.

transformation (Fig. 3.2). This transformation realized unto the grid cells was necessary in order to extend the neighborhood criteria from just contiguity to distance-based neighbors (*k*-nearest neighbors) (Pebesma & Bivand, 2019).

Using *k*-nearest neighbors typically leads to asymmetric neighbors. However, this is not the case as all centroids are uniformly spaced. A key advantage of using distance-based neighbors to ordinary polygon contiguity is that it ensures that all fishnet grid cells polygon representation (centroids) have *k* neighbors. It is common practice to

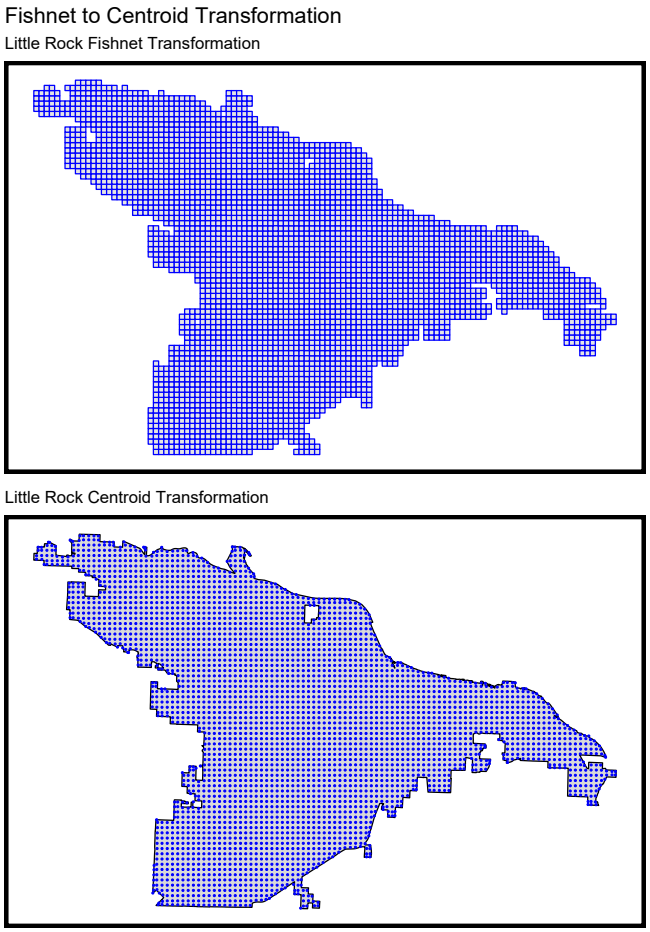

Figure 3.2: Panels showing fishnet grid-cell to centroid transformation representation of Little Rock, AR.

use $k = 8$ or $k = 4$ neighbors which are formally know as "Queen Case" and "Rook Case" for the number of desired neighbor (Figure 3.3). For this research, $k = 8$ nearest neighbors were used and located using the function *knearneigh* and *Knn2nb* from the package **spdep**. Following on the identification of all 8 neighbors for each centroid, their respective weights were assign using the function *nb2listw* from the package **spdep**.

As an example, consider Fig. 3.3 ("8 Nearest Neighbors") as a zoomed in portion of Fig. 3.2. The numbers represent the fishnet *Grid ID*, in essence *Grid ID 1* will have the following list of neighbors *2,3,4,5,6,7,8,9*. The following step after the identification of the neighbors of *Grid 1* is to assignment spatial weights to the list of neighbors. As the term

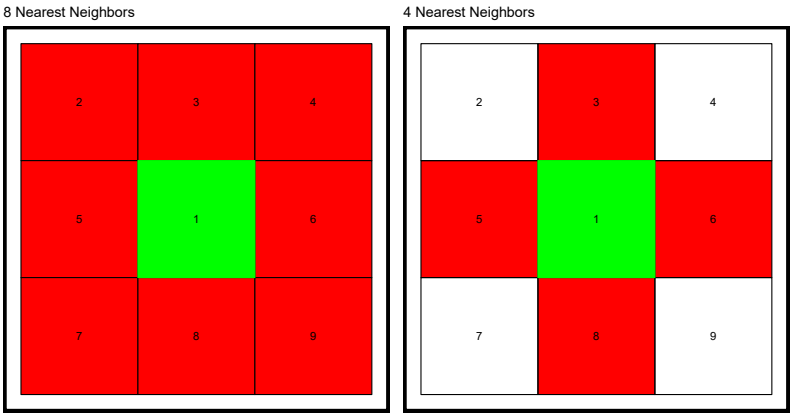

Figure 3.3: Fishnet grid cell representation of Queen case and Rook case neighborhood definition, $k = 8$ and $k = 4$ respectively.

weight implies, it is how much value we want to extract from each neighbors. Assigning equal weights to each grid's neighbors list methodology was used in this research, this suggest that each neighbor will have a corresponding weight of $\frac{1}{8}$. This weight is then used to compute the mean neighbor values as $weight = \frac{1}{8}\sum_{i=2}^{9} weight\ for\ neighbor_i$. This is equivalent to summing all eight mental health incident case that fell within the neighbor grid cell then dividing by 8. Having obtained both neighbors and their respective weights, the following step was to test for the presence of spatial auto-correlation using both Global Moran's I and Local Moran's I.

Global Moran's I

The process for calculating the global test for spatial auto correlation uses local relationships between the observed spatial entity value and its defining neighbors (Bivand et al., 2008).

Definition 3.1.1 (Global Moran's I)**.** Let y_i be the i^{th} observation, with the mean being $\bar{y}$, and let and w_{ij} be the spatial weight of the link between i and j, then Global Moran's I

statistic is given by the following formula:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})},$$

where I represents the ratio of the product of the variable of interest, and adjusted for the spatial weights used.

Centering on the mean is equivalent to asserting that the correct model has a constant mean, and that any remaining patterning after centering is caused by the spatial relationships encoded in the spatial weights.

Local Moran's I

Localized tests are built by breaking global measures into components which aids in the detection of clusters and hot-spots, where, clusters are defined as groups of observations where neighbors have similar features and hot-spots are groups of observations with distinct neighbors. (Bivand et al., 2008).

Definition 3.1.2 (Local Moran's I)**.** Local Moran's I_i values consist of the n individual components added to produce the global Moran's I (definition3.1.1): where the assumption is that the global mean $\bar{y}$ is an accurate summary of the variable of interest y. Note that here we do not center the two components in the numerator, $(y_i - \bar{y})$ and $\sum_{j=1}^{n} w_{ij}(y_j - \bar{y})$.

$$I_i = \frac{(y_i - \bar{y})\sum_{j=1}^{n} w_{ij}(y_j - \bar{y})}{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}.$$

The global Moran's I computed using the function *moran.test* from the **spdep** R packaged produced a single value of 0.22923. We note here that the the global Moran's I has an alternative representation as the slope of the Ordinary Least Square (OLS) regression line in the Fig. 3.4 describing the universal spatial autocorrelation of the data.

To test the significance of Global Moran's I statistic, a permutation bootstrap test with

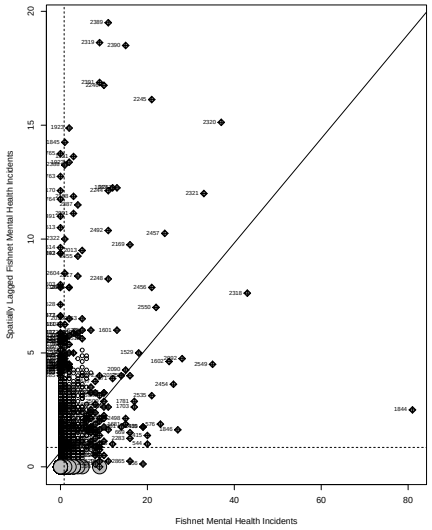

Figure 3.4: Moran's Plot described by the slope of the OLS regression between fishnet mental health incidents and spatially lagged fishnet mental health incidents.

999 simulations was conducted via the *moran.mc* function from the **spdep R** package. The permutation test produced a sampling distribution of the test statistic Moran's I under the null hypothesis of no spatial auto-correlation, which was used to derive a (pseudo) p-value. The (pseudo) p-value of a permutation test is computed using the following formula:

$$\frac{N_{extreme} + 1}{N + 1},$$

where $N_{extreme}$ represent the number of simulated Moran's I values more extreme than the observed Moran's I statistic and N denotes the total number of simulations (gimond2019). The sampling distribution and the observed value of Moran's I is shown on 3.5 for a visual illustration of this test.

Note that the observed value of the Global Moran's I statistic was the max when compared to the simulated values obtained from the permutation test. These results provide a pseudo p-value of 1/1000 = 0.001 indicating that there is a 0.1% probability of observing a test statistic that is as or more extreme compared to the current observed value of the Moran's I under the null hypothesis H_0. With the statistical significance level

Figure 3.5: Sampling Distribution and Observed Value for the Moran's I test statistics.

of the Global Moran's I statistics value established, a localized Moran's test was conducted to identify the location of the possible mental health incident clustering using the function *localmoran* from the **spdep** package. Similar to the global Moran's I described above, the local Moran's I evaluates the level of spatial auto-correlation among the k-nearest fishnet grid cells ($k = 8$, here) surrounding a given fishnet grid cell. Local Moran's test also computes the (pseudo) P-value indicating the significance of the spatial auto-correlation at the level of each fishnet grid cell. Using a significance level of $\alpha = 0.05$ to determine which grid cells indicate a significant level of clustering will be flawed as the local Moran's test executes multiple comparison test (Anselin, 1995). To address the multiple comparison test issue, a Bonferroni p-value adjustment was implemented using the function *p.adjustSP* from the **spdep** package thereby allowing the use of $\alpha = 0.05$ to determine significance post p-value adjustment . For the following figure, the first panel shows the count of Health Incident events; Panel 2 shows the local Moran's I Statistic value at each grid cell, the final panel shows areas that exhibit statistical significant clustering (Gimond, 2019).

The presence of positive and significant spatial auto-correlation in the mental health incidents data clearly substantiates our claim that such events are clustered in space,

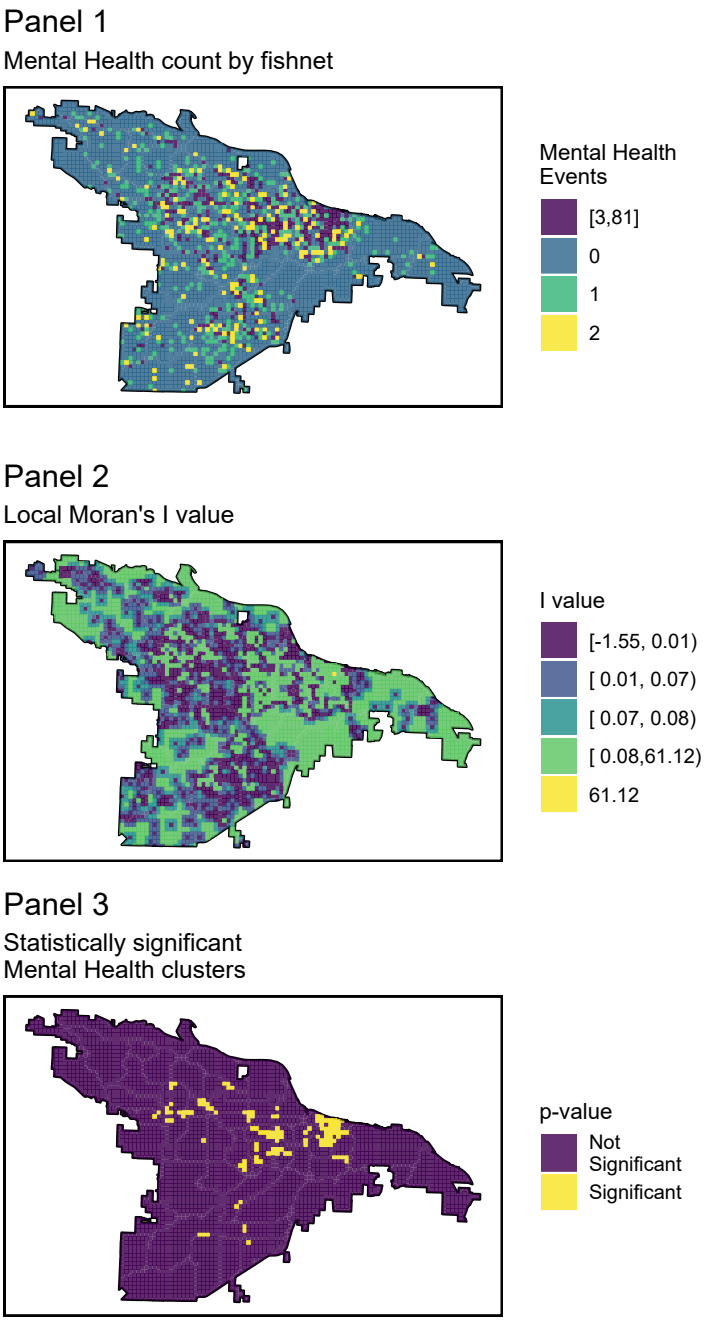

Figure 3.6: Local Moran's I plot illustrating the spatial clusters of mental health incident calls in Little Rock, AR.

instead of uniformly distributed over the entire region of interest. Having obtained such results is essentially the first step in the process of identifying a proper model (Pebesma & Bivand, 2019).

3.1.2 Performance Comparison

Table 3.1: Model Performance Comparison

	MAPE Mean	MAPE SD	MAE Mean	MAE SD	RMSE Mean	RMSE SD
Poisson GLM	1.3112	0.0308	0.9098	0.2699	2.9166	1.5893
Random Forest	1.306	0.0346	0.8677	0.1708	2.1904	0.9008
Manski Model	1.302	NA	0.7708	NA	2.5832	NA
Spatial Durbin	1.316	NA	0.6356	NA	2.135	NA

We compare the predictive performance of the four candidate methods on Table 3.1, and report the mean and standard deviation for each error measure. To better assess the accuracy of the models, we use four different error measures: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) & Root Mean Square Error (RMSE), see Table 3.1 above. The errors were calculated in a supervised learning set-up, where both Poisson regression and Random Forest models were built using leave-one-group-out cross-validation with the number of folds being equal to five. Below, we define the different error measures used to compare and describe the best performing model according to that criterion.

First, the Mean Absolute percentage Error (MAPE) statistic captures the model's accuracy in terms of percentage error. The MAPE is calculated using the following formula:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{A_i - F_i}{A_i}| \times 100,$$

where A_i is the i^{th} actual observation and F_i is the i^{th} forecast value. Since the MAPE expresses the error as percentage, it can be relatively easier to interpret when compared to other statistic measures. The lower the percentage error, the more accurate the model represents the data. For a given model, it can be concluded that on average, the forecast is off by the MAPE. We can clearly see that on average all models forecasts were off by approximately 1.3% with a standard deviation of approximately 0.0308 and 0.0346 for the

Poisson GLM and Random Forest respectively. In terms of MAPE, all models perform relatively the same with the Manski model having the smallest.

The Mean Absolute Error (MAE) statistic captures on average how large the forecast error is expected. The MAE is given by the formula

$$MAE = \frac{\sum_{i=1}^{n} |A_i - F_i|}{n},$$

where A_i is the i^{th} actual observation and F_i is the i^{th} forecast value. Spatial Durbin error model had on average the smallest forecast error of 0.6356 followed by the Manski Model with a MAE of 0.7708 and Poisson GLM having the largest forecast error of 0.9098.

The Root Mean Square Error (RMSE) or otherwise also known as the Root Mean Square Deviation calculates the square root of the average of the square errors. The RMSE measures the spread of the prediction errors. The RMSE is given by the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (F_i - A_i)^2}{n}}.$$

Spatial Durbin Error model had the smallest RMSE value of 2.135 followed by Random Forest with a RMSE of 2.1904 and the Poisson GLM having the largest RMSE of 2.9166.

3.1.3 Goodness of fit metrics

Table 3.2: Model Goodness of fit Comparision

	R2 Mean	R2 SD	LogDev Mean	LogDev Sd
Poisson glm	0.3927	0.1517	0.6141	0.0509
Random Forest	0.3822	0.0582	0.5844	0.0403
Manski Model	0.4366	NA	0.6124	NA
Spatial Durbin	0.4735	NA	0.7102	NA

In terms of Goodness of fit metrics, the R squared (R^2) values and logarithmic deviance score were used to evaluate the models. The most common measure is perhaps

the R^2 that represents the percentage of variation explained by the model,

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},\ \hat{y}_i = \text{ predicted value of } y_i,\ \bar{y} = \text{ grand mean},$$

thus a larger R^2 is indicative of a better model fit. Note that the Adjusted R^2 value was not computed as it is rather difficult to compute for random forest and use in goodness-of-fit comparison. The Logarithmic deviance Score is a measure of the deviance between the predicted and observed counts, via the log likelihood ratio. To measure this, we calculate the likelihood ratio of the observed value and the predicted value based on a Poisson distribution. The goodness of fit reported here is the negative log of the probability density so a lower value indicates a better predictive ability. As seen in table 3.2, Spatial Durbin error model obtained the largest R square value followed by the Manski Model. Note that despite having obtained the largest R square value *i.e* the best model in terms of R square goodness of fit metric, it obtained the largest logarithmic deviance score thus the worst model in logarithmic deviance score goodness of fit metric for the mental health dataset. Continuing on table 3.2, for the Logarithmic Deviance score goodness of fit metric, the Random Forest model obtained the smallest score. This suggest that Random Forest had the smallest deviance between predicted and observed count of mental health incidents *i.e.* the best model of such category.

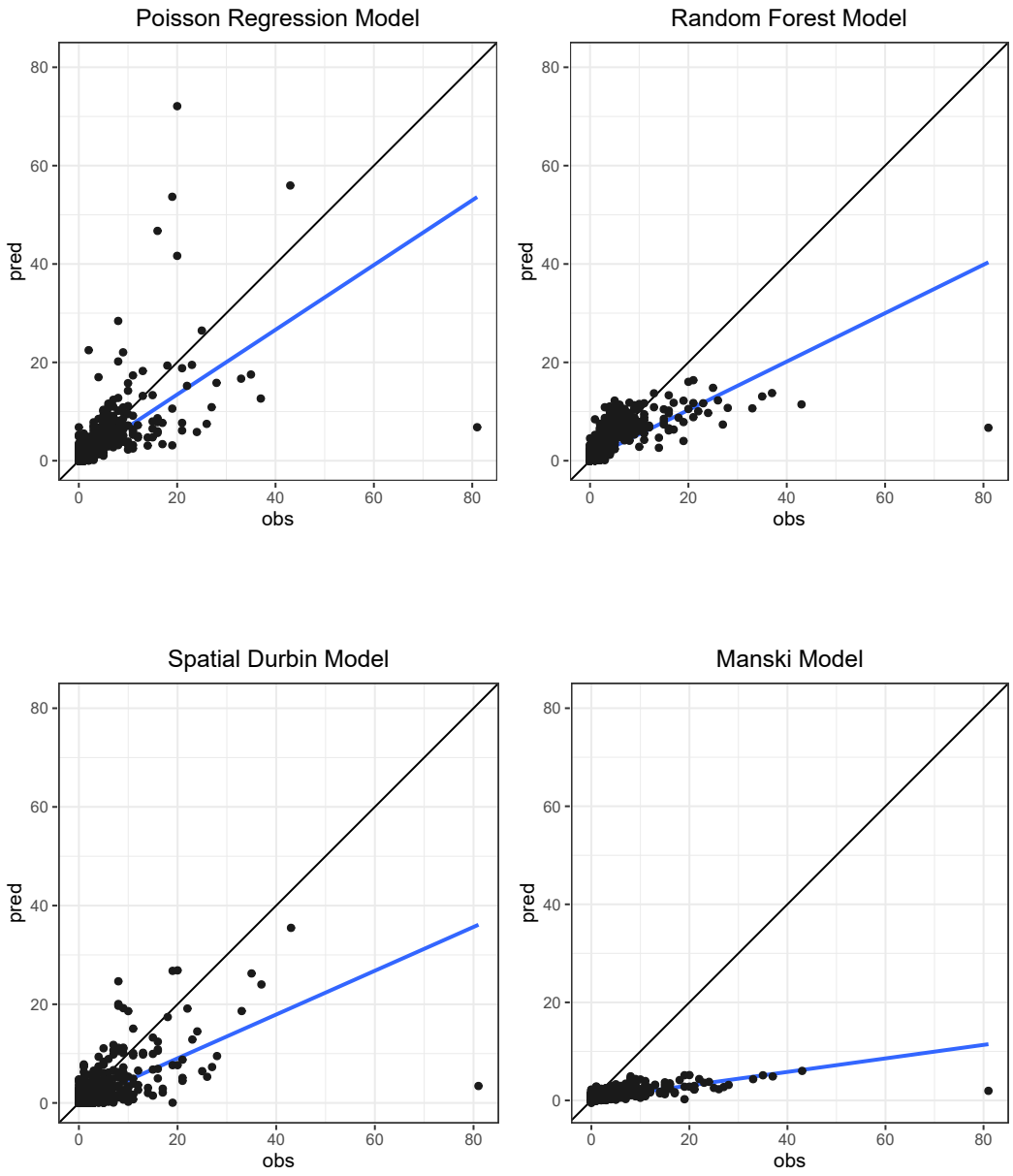

Figure 3.7: Predicted versus observed mental health incident cases plots by the candidate models.

3.1.4 Feature Importance Comparison

Finally, we look at the important features or variables driving the prediction for each of the four candidate methods. We call these measures 'variable importance' following the

nomenclature used by random forest literature, but for purely statistical models such as Poisson regression or spatial Durbin models, the quantities being compared are a measure of each variable's significance. As discussed before, this a key step in the prediction process as the important variables help us in identifying which environmental and social features are predominantly occupying each of these predictive processes, investigate whether they play a risky or protective role and then allocate resources accordingly.

A note about nomenclature for the features plotted on the following figures. There are three unique prefixes linked with each type of feature. Nearest neighbor ('NN') refers to features obtained by calculating the average distance between a fishnet grid cell centroid and its nearest neighbor in Queen case definition. Euclidean distance ('ed') refers to features obtained by calculating the euclidean distance between a fishnet grid cell centroid and its first nearest neighbor. agg refers to the count of mental health incident in a given fishnet grid cell. The term 'agg' was coined based on the *aggregate* function used in R to obtain the count of cases associated per fishnet cell.

Table 3.3: Top ten covariates with decreasing order of significance for each model.

Poisson_GLM	Random_Forest
agg_Rentals_Apts_Over100units	NN_PoliceFacilities
agg_Rentals_Apts_LessThan100units	NN_Banks
agg_MajorDeptRetailDiscount	agg_BusStops
agg_FastFoodAndBeverage	agg_GasStationAndConvMart
agg_MixedDrink_BarRestClub	agg_FastFoodAndBeverage
agg_BusStops	NN_ChildCareServices
NN_ReligiousOrgs	NN_BarberAndBeautyShops
agg_LiquorStores	NN_ChildYouthServices
agg_GasStationAndConvMart	agg_LiquorStores
NN_Unsafe_Vacant_BldgsNEW	NN_ReligiousOrgs

Table 3.3 Continued.

Spatial_Durbin	Manski
agg_Rentals_Apts_Over100units	agg_Rentals_Apts_Over100units
agg_FastFoodAndBeverage	agg_FastFoodAndBeverage
agg_BusStops	agg_BusStops
agg_Rentals_Apts_LessThan100units	agg_GasStationAndConvMart
agg_GasStationAndConvMart	agg_MajorDeptRetailDiscount
agg_MajorDeptRetailDiscount	agg_Rentals_Apts_LessThan100units
agg_HotelMotel.x	agg_HotelMotel.x
agg_MixedDrink_BarRestClub	agg_MixedDrink_BarRestClub
NN_Unsafe_Vacant_BldgsNEW	NN_Unsafe_Vacant_BldgsNEW
agg_LiquorStores	agg_LiquorStores

Table 3.3 summarizes the top ten most influential features from each model. We note here that four similar features were found among the set of top features selected by four models. These common features were: `agg_FastFoodAndBeverage`, `agg_BusStops`, `agg_LiquorStores`, `agg_GasStationAndConvMart`. As the four models highlight the importance of the influence these features had on the models, further interdisciplinary study involving experts from criminology and local law enforcement is required to understand whether any causal relationship exists between these environmental factors and mental health incidents in Little Rock, AR.

The following plots illustrate the feature importance in descending order with respect to each model. In order to create a visual feature comparison between the Random Forest feature importance plot and the remaining models, the $-\log_{10} P$-values of each predictor for each other models were plotted.

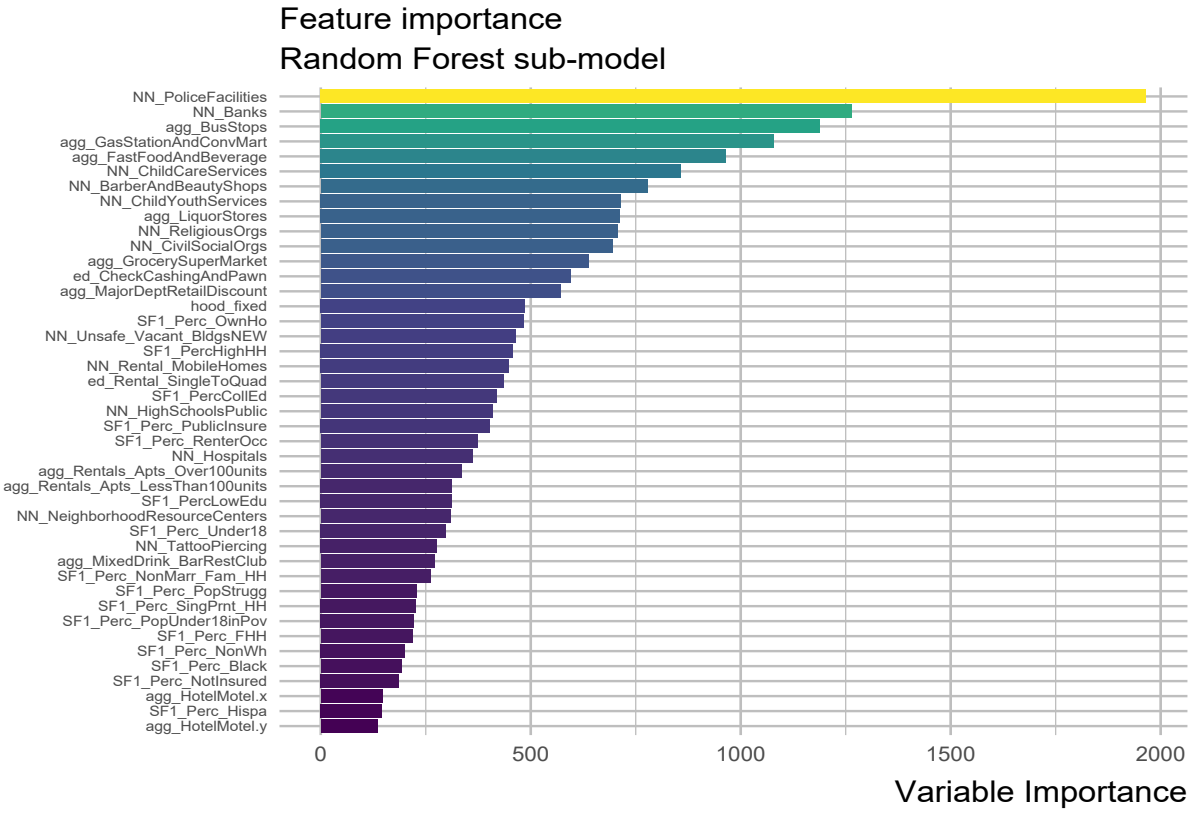

Figure 3.8: Variable Importance for Random Forest.

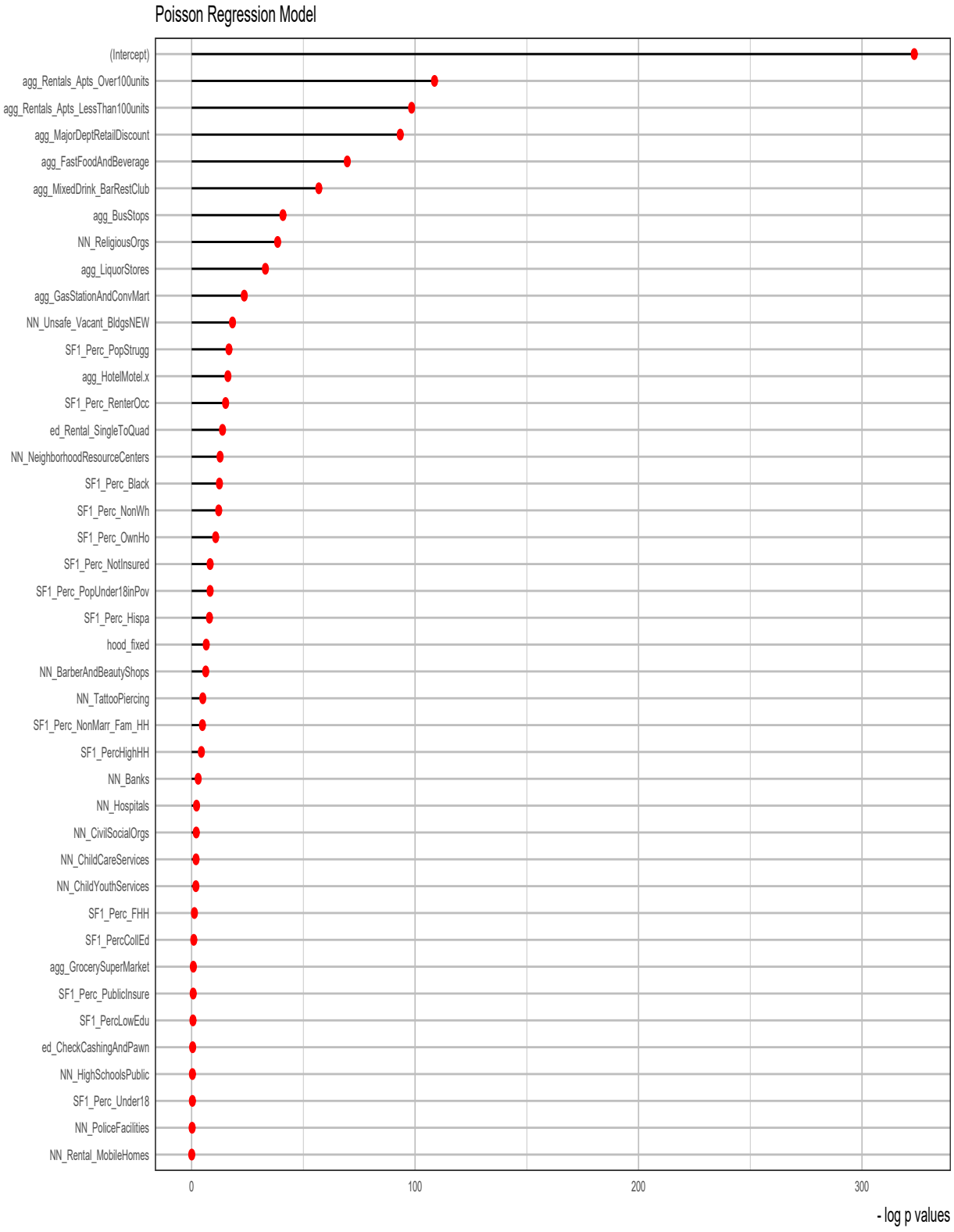

Figure 3.9: Poisson Regression variables in decreasing order of significance via - log P-values.

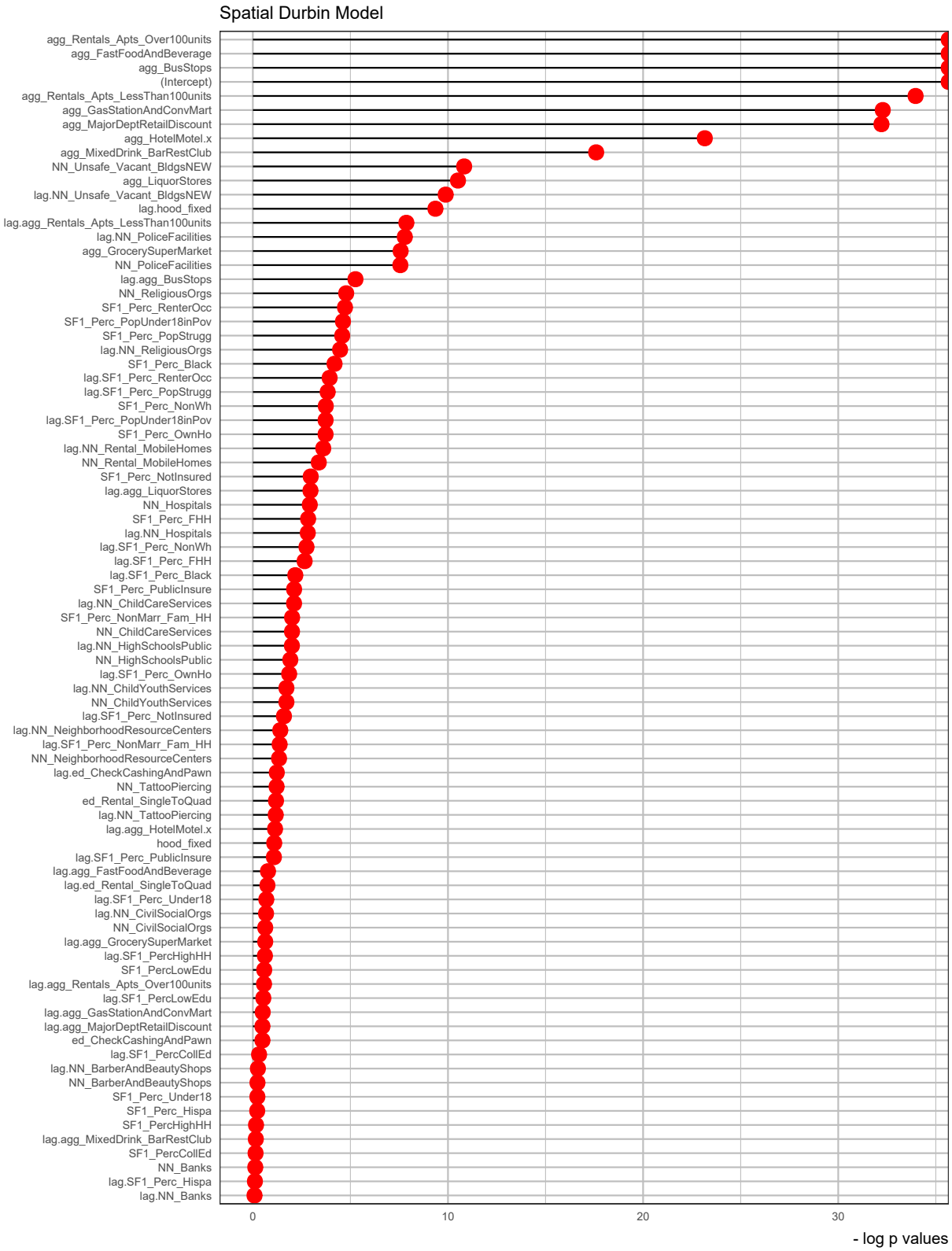

Figure 3.10: Spatial Durbin Regression variables in decreasing order of significance via -log P-values.

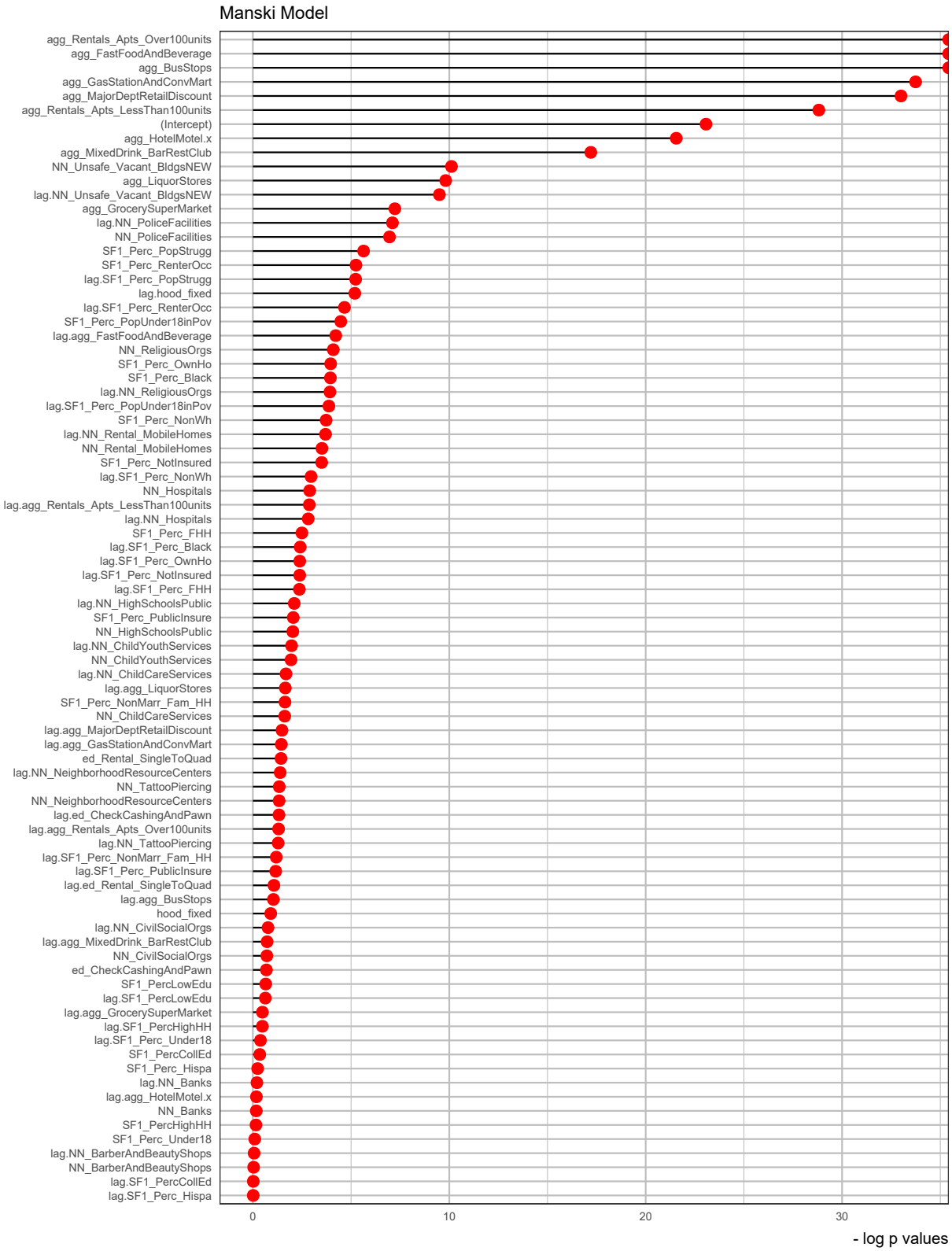

Figure 3.11: Manski Regression variables in decreasing order of significance via - log P-values.

Chapter 4

Conclusion

In this thesis we used a machine learning framework to understand the effect of socio-demographics as well as environmental factors in predicting the spatial clusters of mental health incidents in Little Rock, Ar. The use of spatial auto-correlation Moran's I as exploratory data analysis revealed an uneven distribution of mental health incidents across the areas of study. The primary aim of this thesis was to expand the methodology under risk terrain management by incorporating statistical models to predict mental health incidents based on socio-economic predictors and environmental factors. We compared four different statistical methods prediction accuracy and goodness of fit to provide insight on the list of factors affecting mental health incidents in Little Rock, Ar. Results indicate that in terms prediction accuracy, the spatial econometric models (Manski and Spatial Durbin Error model) performed better than their models counter parts by a small margin. For goodness of fit test R squared and Logarithmic deviance score respectively, Spatial Durbin error model and Random Forest model performed the best. The incorporation of these models under the risk terrain management would definitely serve law enforcement agencies to properly allocate resources to address the unequal distribution of theses incidents.

Furthermore, if law enforcement agencies adopt this framework, creating a meta model from the models generated would serve as a better tool if indecisive of which model to select based on prediction accuracy or goodness of fit. In addition to creating a meta model, the implementation of temporal features and regularization parameters would provide if not better prediction and model goodness of fit results. Finally in would be meaningful to determine how these associations or patterns change in relation to a post Covid-19 pandemic.

Bibliography

ANSELIN, L. (1995). Local indicators of spatial association—lisa. *Geographical Analysis* **27**, 93–115.

BIVAND, R. & WONG, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST* **27**, 716–748.

BIVAND, R. S., PEBESMA, E. J., GOMEZ-RUBIO, V. & PEBESMA, E. J. (2008). *Applied spatial data analysis with R*, vol. 747248717. Springer.

BOOTS, B. (2001). Spatial pattern, analysis of. In *International Encyclopedia of the Social & Behavioral Sciences*, N. J. Smelser & P. B. Baltes, eds. Oxford: Pergamon, pp. 14818 – 14822.

BREIMAN, L. (2001). Random forests. *Machine learning* **45**, 5–32.

CAPLAN, J. M., KENNEDY, L. W., BARNUM, J. D. & PIZA, E. L. (2015). Risk terrain modeling for spatial risk assessment. *Cityscape* **17**, 7–16.

CAPLAN, J. M., KENNEDY, L. W. & MILLER, J. (2011). Risk terrain modeling: Brokering criminological theory and gis methods for crime forecasting. *Justice Quarterly* **28**, 360–381.

DE JONG, P., SPRENGER, C. & VEEN, F. (1984). On extreme values of moran's i and geary's c (spatial autocorrelation). *Geographical Analysis* **16**, 17–24.

DRAWVE, G. (2016). A metric comparison of predictive hot spot techniques and rtm. *Justice Quarterly* **33**, 369–397.

ELHORST, J. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Springer.

GIMOND, M. (2019). Intro to GIS and Spatial Analysis.

GOTWAY, C. A. & STROUP, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics* , 157–178.

HALLECK VEGA, S. & ELHORST, J. P. (2015). The slx model. *Journal of Regional Science* **55**, 339–363.

JAMES, G., WITTEN, D., HASTIE, T. & TIBSHIRANI, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

LUM, K. & ISAAC, W. (2016). To predict and serve? *Significance* **13**, 14–19.

MOHLER, G. et al. (2013). Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics* **7**, 1525–1539.

MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. & TITA, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association* **106**, 100–108.

MOHLER, G. O., SHORT, M. B., MALINOWSKI, S., JOHNSON, M., TITA, G. E., BERTOZZI, A. L. & BRANTINGHAM, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American statistical association* **110**, 1399–1411.

MONTGOMERY, D. C., PECK, E. A. & VINING, G. G. (2006). *Introduction to Linear Regression Analysis (4th ed.)*. Wiley & Sons.

PEBESMA, E. J. & BIVAND, R. (2019). Spatial Data Science.

PERRY, W. L., MCINNIS, B., PRICE, C. C., SMITH, S. & HOLLYWOOD, J. S. (2013). Predictive policing: Forecasting crime for law enforcement .

TOWNSLEY, M., HOMEL, R. & CHASELING, J. (2000). Repeat burglary victimisation: Spatial and temporal patterns. *Australian & New Zealand journal of criminology* **33**, 37–63.

WANG, X. & BROWN, D. E. (2012). The spatio-temporal modeling for criminal incidents. *Security Informatics* **1**, 2.