Statistical analysis of commuting patterns in and out of Fort Wayne, IN and Cook County, IL

Leah Morris*, Megan O’Sullivan*,and LaDavie’a Shears*

Department of Mathematical Sciences*, Purdue University Fort Wayne

Supervisor: Alessandro Maria Selvitella*

Introduction

Several models have been developed to study the complex movement patterns of human migration [5] , trade [2], telecommunication [6], and other mobility networks. Since its development in the mid-20th century, the gravity model [8] has been the predominant model used to forecast movements resulting from a variety of human activity from the individual [7] to the international level [1]. It predicts travel between origin and destination cities using the populations of each and the distance traveled, similar to Newton’s gravity law [3].

This project aims to examine the commuting flow of residents of the Fort Wayne, Indiana area, and Chicago, Illinois residents using the gravity model as presented in the article *A universal model for mobility and migration patterns*. We will examine results from various regression models. The commuting flow of Fort Wayne, Indiana will be defined as commuters within a 100 mile radius around Allen County. Similarly, the commuting flow of Chicago, Illinois will be defined as commuters within a 100 mile radius of Cook County.

The Gravity Model

Concerning human mobility, the gravity model can be used to predict migration between locations. Suppose that T_{ij} is the number of individuals that move between locations i and j, m_i the source population, n_j destination population, and r_{ij} the distance between locations i and j. Here α and β represent parameters to be determined and weight the dependence of T_{ij} on the two population sizes. The *deterrence function*, $f(r_{ij})$ can take on different forms such as the power function, $f(r) = r^\gamma$, or the exponential function, $f(r) = e^{dr}$.
The gravity model is then given by

$$T_{ij} = \frac{m_i^\alpha n_j^\beta}{f(r_{ij})}.$$

or equivalently, by taking the log of both sides $\ln(T_{ij}) = \ln(\frac{m_i^\alpha n_j^\beta}{f(r_{ij})})$, we get

$$\ln(T_{ij}) = \alpha \ln(m_i) + \beta ln(n_j) - \ln(f(r_{ij}))$$

which is linear in the "log"-variables. If we choose r^γ for the *deterrence function* with r being the distance between two locations, i and j, it becomes more evident how to use a linear regression to fit the model to data.

$$\ln(T_{ij}) = \alpha \ln(m_i) + \beta ln(n_j) - \gamma ln(r_{ij}),$$

Gravity Model obtained from: *A universal model for mobility and migration patterns*

Consider the exponential function as the *deterrence function*, $f(r_{ij})$= $e^{dr_{ij}}$. In our formula we substitute $f(r_{ij})$ for $e^{dr_{ij}}$ and get

$$\ln(T_{ij}) = \alpha \ln(m_i) + \beta ln(n_j) - \ln(e^{dr_{ij}}),$$

we know $ln(e) = 1$, so we rewrite the model as:

$$\ln(T_{ij}) = \alpha \ln(m_i) + \beta ln(n_j) - dr_{ij},$$

where, as mentioned, r_{ij} is the distance between i and j and our coefficients are α, β, and d.

Methods

Using the method of ordinary least squares, we will fit the gravity model to our data. In doing so, we will find the coefficients α, β, and γ which minimize the sum of the squared differences between the data provided (observed) and predicted by the model. These estimators will inform us on how each explanatory variable impacts the commuting flow between two cities. When each predictor is considered in isolation by holding all others fixed, these coefficients give us the slope representing the magnitude and direction of the relationship between x_i and y.
We will then look at the coefficient of determination, R^2, the proportion of the given data which is represented by the model. If our R^2 value is close to 1, we know the model has a good fit to the data and can be considered a good model. [4].

$$R^2 = 1 - \frac{SSR}{SST}$$

Given that the gravity model hinges on population as a determining factor of the linear fit, we conjectured that community data with a larger urban population might be better predicted by this model, since in physics, gravitational force increases with larger objects. We limited the scope of our data collection to distances 100 miles or less, since our intention was to capture daily commuter travel patterns. As a result, available data about commuting to distant states is not included in this version of our investigation.

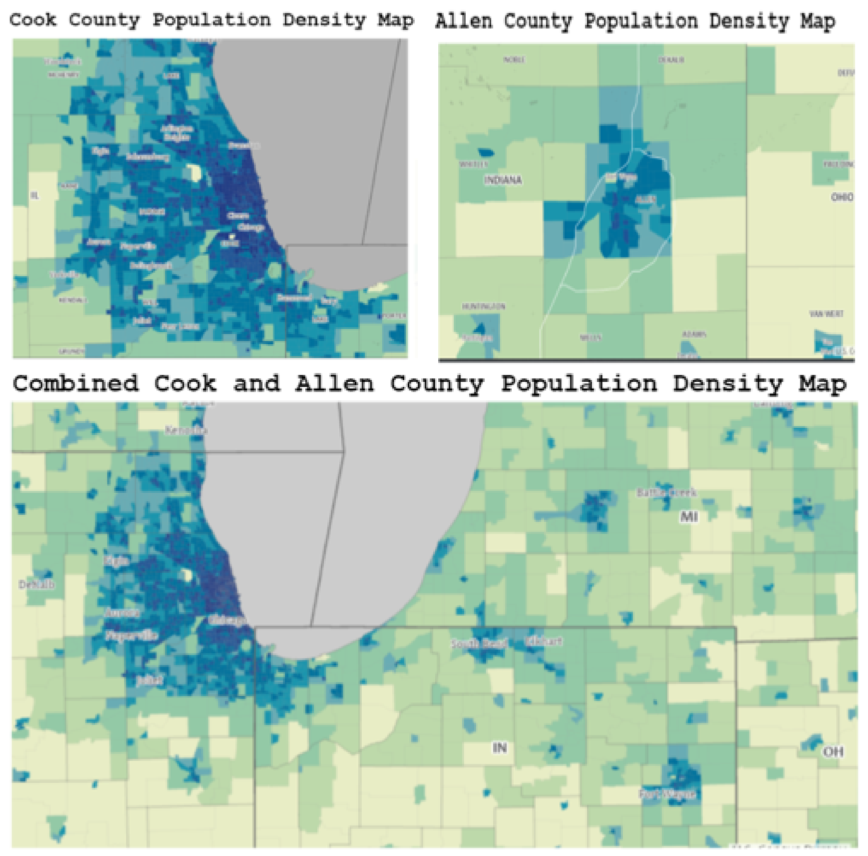

Fig. 1: Population Density Map of Allen and Cook County

Results

We found the gravity model using the power function to be most effective in modeling the data for Cook County alone, followed closely by aggregating the Allen and Cook County data and examining it together. The Allen County data was moderately well captured by the gravity model. This pattern held true for the gravity model using the exponential function as well. The coefficients of determination for all three cases were extremely close between the various models.

County Linear Model	Coefficient of Determination
Cook County (Power)	0.8856
Cook County (Exponential)	0.8856
Cook and Allen County (Power)	0.8596
Cook and Allen County (Exponential)	0.8345
Allen County (Power)	0.7622
Allen County (Exponential)	0.7230

Considerations and Conclusions

As we had predicted, due to the nature of the gravity model being analogous with phenomena seen in Newtonian physics, we saw greater success when applying the model to areas with larger populations and greater population density, notably Cook County alone, or Cook County combined with Allen County. We have included population density maps that illustrate the differences between the two regions.

Additional factors outside of population and population density that may be contributing to the differences in R^2 values between the regressions for Cook and Allen could be availability of transportation infrastructure; Cook County and the surrounding areas have significantly more options for commuters wishing to travel for work, including city buses, trains, commuter trains, and expressways. Further investigations into the modes of transportation elected by commuters could help illuminate whether transportation options have a significant impact on the effectiveness of the gravity model in predicting commuting patterns for these areas.

Additionally, because the gravity model does not account for direction of travel, but only the populations of the counties traveled and the distance between them, we were not able to evaluate differences between commuting out of versus commuting into Allen or Cook counties. Further, a limitation of the gravity model in predicting commuting information specifically is that it cannot account for commuting within one county, since in Newtonian physics, an object cannot exert gravitational pull on itself. Also, the data was aggregated by county, rather than by raw commuter travel distance, so this presented limitations in evaluating intra-county travel. Alternate models may be useful to mitigate these limitations, particularly the radiation model, which has been used extensively to capture complex movement patterns.

Contributions

All authors were involved and informed through each step of this project. LaDavie’a Shears contributed in data research, document formatting, coding in R, and analysis. Megan O’Sullivan contributed in literature review, data research, data organization, and formally writing the interpretation of results. Leah Morris contributed in data research, coding in R, analysis, and final revisions.

References

[1] D Balcan et al. “Multiscale mobility networks and the spatial spreading of infectious diseases”. In: *Proceedings of the National Academy of Sciences* 106(51) (2009), pp. 21484–21489.

[2] K Bhattacharya et al. “The International Trade Network: weighted network analysis and modelling”. In: *Journal of Statistical Mechanics* 453 (7196) (2008), P02002.

[3] Stewart Erlander S. *The gravity model in transportation analysis: theory and extensions.* Vol. 1. VSP, 1990.

[4] Edward W. Frees. *Regression Modeling with Actuarial and Financial Applications.* Vol. 1. Cambridge University Press, 2009.

[5] Baraba´ si AL Gonzalez MC Hidalgo CA. “Understanding individual human mobility patterns”. In: *Nature* 453(7196) (2008), pp. 779–782.

[6] G Krings et al. “Urban gravity: a model for inter-city telecommunication flows”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2009(07) (2009), p. L07003.

[7] Simini F Pappalardo L. “Data-driven generation of spatio-temporal routines in human mobility”. In: *Data Mining and Knowledge Discovery* 32(3) (2018), pp. 787–829.

[8] GK. Zipf. “The P1 P2 / D Hypothesis: On the Intercity Movement of Persons.” In: *American Sociological Review* 11(6) (1946), pp. 677–686.