Food Deserts or Food Oases? Predicting Grocery Store Locations in Hamilton, Ontario

Zehui Yin[a,*]

[a] *School of Earth, Environment & Society, McMaster University, 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada*

Abstract

This is the abstract.

It consists of two paragraphs.

Keywords: Grocery stores, Food environments, Hamilton

1. Introduction and Background

Food is a necessity for human beings. In urban areas or large cities, residents typically rely on grocery stores for their daily food needs. Geographic access to grocery stores and affordable food plays an important role in promoting a healthy diet and has implications for personal health (Caspi et al., 2012; Minaker, 2016; Kirkpatrick et al., 2014). The ease of accessing grocery stores and obtaining the needed food is thus an important topic for both public health and urban planning.

There is extensive literature on food access that examines this topic (Christian, 2012; Widener et al., 2015; Farber et al., 2014; Widener et al., 2017). These analyses mostly focus on the demand side, where consumers navigate and choose which retailers to purchase from. However, a gap exists in studying how retailers choose the locations of their stores to serve the market or the demand.

In this paper, I utilize open-source data from the 2021 Canadian Census (Statistics Canada, 2022) and OpenStreetMap (OpenStreetMap contributors, 2017) to examine the spatial pattern of grocery store locations, using Hamilton, Ontario, as a case study. The research question is: What areas in Hamilton have more or fewer grocery stores compared to other areas?

2. Data and Methods

2.1. Study Area

My study area is the City of Hamilton, located on the west side of Lake Ontario in the province of Ontario. It has a population of 569353 according to the 2021 Canadian census (Statistics Canada, 2022). The Niagara Escarpment runs through the middle of the city, dividing it into two parts. Figure 1 below shows the study area.

*Corresponding author

Email address: `yinz39@mcmaster.ca` (Zehui Yin)

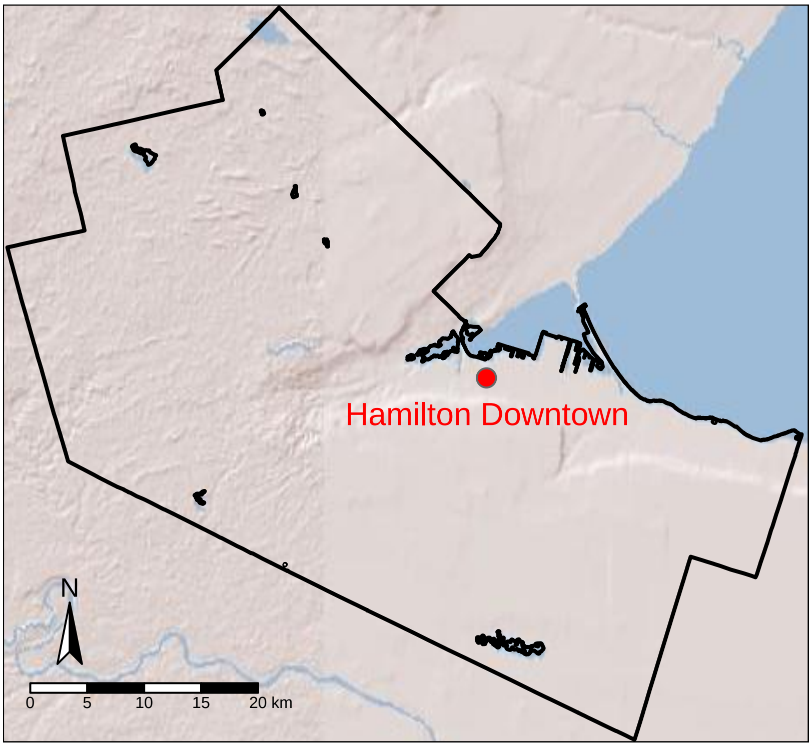

Figure 1: Study Area: Hamilton, Ontario

2.2. *Data Sources*

I utilize three data sources in this paper, as shown in the Table 1 below. I gathered grocery store locations in Hamilton from OpenStreetMap (OpenStreetMap contributors, 2017) via the Overpass API (OpenStreetMap Wiki, 2024). The Hamilton Street Railway (HSR) Fall 2024 GTFS static data were downloaded from Open Hamilton (City of Hamilton, 2024). The dissemination area spatial data and 2021 Canadian census variables were obtained from Statistics Canada (2022) through the cancensus package in R (von Bergmann et al., 2022).

Name	Source	URL	Accessed Date
Grocery Stores in Hamilton	OpenStreetMap contributors (2017)	`https://overpass-turbo.eu/index.html`	2024-10-04
HSR Fall 2024 GTFS Static	City of Hamilton (2024)	`https://opendata.hamilton.ca/GTFS-Static/`	2024-10-04
Dissemination Area and Census Data in Hamilton	Statistics Canada (2022)	`https://censusmapper.ca/api`	2024-11-16

Table 1: Data Sources

To facilitate reproducibility and open science, all the data used in this paper have been packaged into an R package (Yin, 2024), which is hosted on GitHub. You can access it at `https://github.com/zehuiyin/geog712package`.

2.3. Methodology

The grocery store locations in Hamilton were intersected and aggregated to the census dissemination areas. The dependent variable of interest is the count of grocery stores in each dissemination area in Hamilton. There are 891 dissemination areas in Hamilton; however, due to data missingness, only 876 of them are used in the regression analysis. A considerable portion of the dissemination areas in our sample do not contain any grocery stores. Therefore, standard count models such as Poisson regression would be invalid due to the excessive zero values. To model this variable of interest, I fit a hurdle model and a zero-inflated negative binomial regression model with spatially lagged dependent variables.

The zero-inflated negative binomial regression model follows Equations 1 and 2 below (NCSS Statistical Software, n.d.). The variable specification in the hurdle model is exactly the same as in the zero-inflated negative binomial regression model. The hurdle model is set up with the zero component as a binomial logit model and the count component as a truncated negative binomial logit model. I decided to use negative binomial regression instead of Poisson regression because the count distribution is highly skewed, making it unlikely that the mean and variance would be the same for my variable of interest. Additionally, due to the large number of zeros in my sample, the zero-inflated regression is preferred, as it can generate zero values from two sources, unlike the hurdle model, which generates zeros from only one source.

$$Pr(GroceryStore_i = j) = \begin{cases} \pi_i + (1-\pi_i)g(GroceryStore_i = 0) & \text{if } j = 0 \\ (1-\pi_i)g(GroceryStore_i) & \text{if } j > 0 \end{cases} \tag{1}$$

$$\begin{aligned} logit(\pi) &= \rho \mathbf{W} GroceryStore + \tilde{\mathbf{x}}\tilde{\boldsymbol{\beta}} \\ \log\{E[g(GroceryStore)]\} &= \rho \mathbf{W} GroceryStore + \mathbf{x}\boldsymbol{\beta} \\ g(GroceryStore_i)&\text{is the negative binomial distribution} \\ &\mathbf{W}\text{: a row-normalized queen contiguity matrix} \end{aligned} \tag{2}$$

3. Results

3.1. Descriptive Statistics

Based on the bar chart in Figure 2, the distribution of the number of grocery stores in dissemination areas in Hamilton is highly skewed, with a large number of dissemination areas having no grocery stores. According to Figure 3, most grocery stores are located near the centre of Hamilton, within the Niagara Escarpment. There are also some dissemination areas at the edge of the city with grocery stores. In the right plot, the number of HSR bus stops per square kilometre is classified into three categories: low (0 to 50th percentile), middle (50th to 75th percentile), and high (75th to 100th percentile). The area with the highest transit service is also in the centre of Hamilton. Beyond the Niagara Escarpment, there are few transit stops.

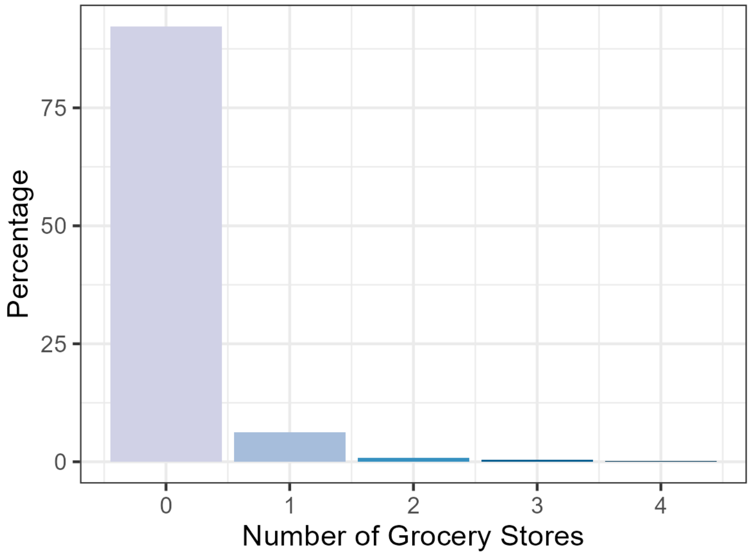

Figure 2: Number of Grocery Stores in Dissemination Areas in Hamilton

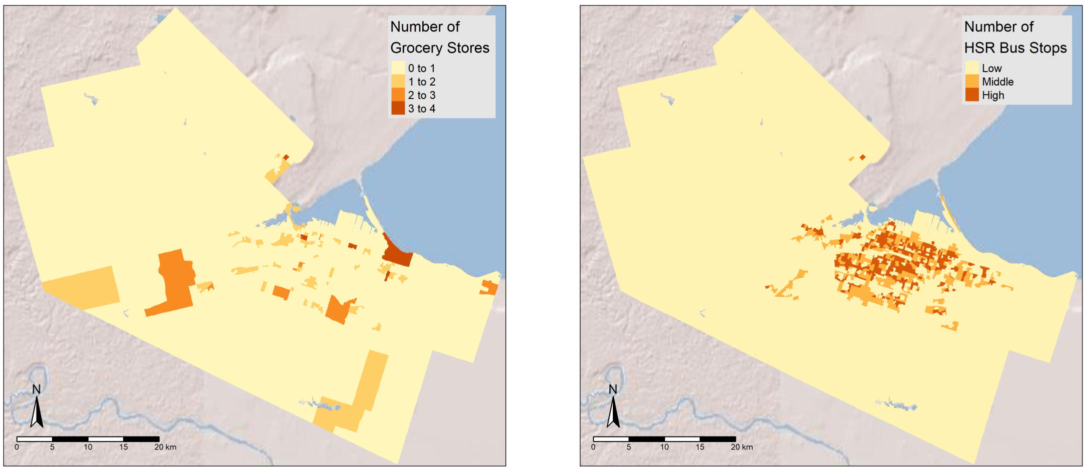

Figure 3: Number of Grocery Stores (left) and HSR Bus Stops (right) at Dissemination Areas in Hamilton, Ontario

4. Regression Results

The hurdle model has a McFadden pseudo R^2 of 0.15, while the zero-inflated negative binomial regression model has a McFadden pseudo R^2 of 0.17. Table 2 below presents the regression results.

	Hurdle model	Zero-inflated model
Count model: Spatial lag of grocery store count	-2.05	-2.99^{***}
	(2.37)	(0.84)
Count model: Percentage of population aged below 24 years old	-0.06	0.03
	(0.11)	(0.04)
Count model: Percentage of population aged above 65 years old	-0.02	0.02
	(0.05)	(0.02)
Count model: Percentage of population don't know official language	-0.11	-0.03
	(0.26)	(0.10)
Count model: Percentage of population don't speak official language at home	0.08	0.17^{**}
	(0.12)	(0.06)
Count model: Percentage of population live in single detached houses	-0.01	$-0.01^{\cdot}$
	(0.02)	(0.01)
Count model: Percentage of population have annual total income less than 40K	0.02	-0.03
	(0.09)	(0.03)
Count model: Percentage of population have annual total income more than 100K	-0.04	-0.02
	(0.10)	(0.04)
Count model: Percentage of population that are married or live in common-law	0.02	0.03
	(0.08)	(0.03)
Count model: Natural log of (population density + 1)	-0.59	-0.38^{*}
	(0.37)	(0.17)
Count model: Natural log of distance from DA centroid to Hamilton downtown	-0.25	-0.50^{**}
	(0.43)	(0.19)
Zero model: Spatial lag of grocery store count	0.02	-8.62^{*}
	(0.58)	(3.38)
Zero model: Percentage of population don't speak official language at home	0.10^{**}	0.14
	(0.03)	(0.11)
Zero model: Percentage of population that are married or live in common-law	-0.04^{*}	0.11
	(0.02)	(0.07)
Zero model: Natural log of (population density + 1)	0.73^{*}	-1.71^{*}
	(0.29)	(0.74)
Zero model: Number of HSR bus stops (50-75 percentile)	2.15^{***}	-2.73^{***}
	(0.43)	(0.72)
Zero model: Number of HSR bus stops (75-100 percentile)	1.55^{**}	-1.50^{*}
	(0.49)	(0.74)
Zero model: Natural log of area size in square kilometres	1.39^{***}	-2.36^{**}
	(0.29)	(0.76)
AIC	522.36	508.91
Log Likelihood	-240.18	-233.45
Num. obs.	876	876

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Table 2: Regression results

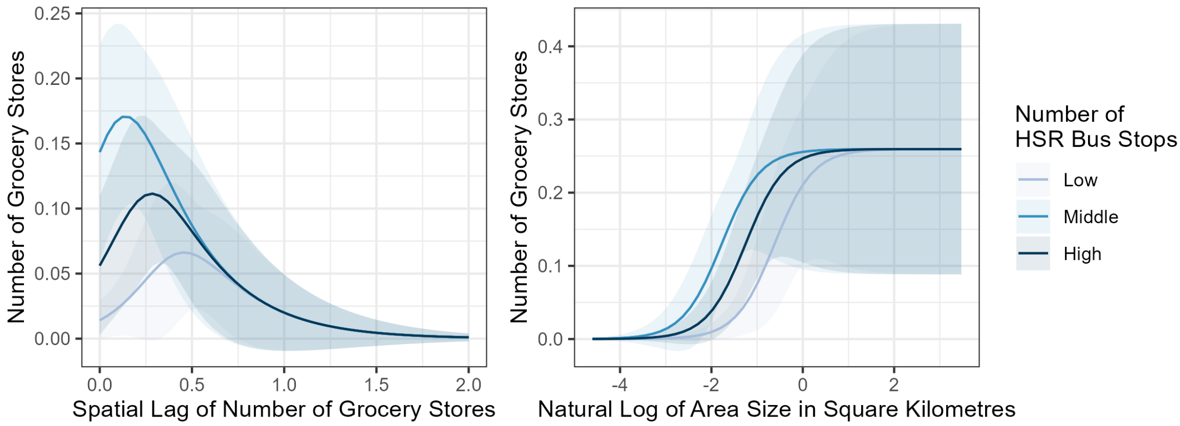

Figure 4: Conditional Prediction Plot for the Zero-inflated Negative Binomial Regression Model

5. Discussion and Conclusion

References

Caspi, C.E., Kawachi, I., Subramanian, S., Adamkiewicz, G., Sorensen, G., 2012. The relationship between diet and perceived and objective access to supermarkets among low-income housing residents. Social science & medicine 75, 1254–1262.

Christian, W.J., 2012. Using geospatial technologies to explore activity-based retail food environments. Spatial and spatio-temporal epidemiology 3, 287–295.

City of Hamilton, 2024. Hsr transit feed. `https://opendata.hamilton.ca/GTFS-Static/`.

Farber, S., Morang, M.Z., Widener, M.J., 2014. Temporal variability in transit-based accessibility to supermarkets. Applied Geography 53, 149–159.

Kirkpatrick, S.I., Reedy, J., Butler, E.N., Dodd, K.W., Subar, A.F., Thompson, F.E., McKinnon, R.A., 2014. Dietary assessment in food environment research: a systematic review. American journal of preventive medicine 46, 94–102.

Minaker, L.M., 2016. Retail food environments in canada: Maximizing the impact of research, policy and practice. Canadian Journal of Public Health 107, eS1–eS3.

NCSS Statistical Software, n.d. Chapter 328 Zero-Inflated Negative Binomial Regression. URL: `https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Zero-Inflated_Negative_Binomial_Regression.pdf`.

OpenStreetMap contributors, 2017. Planet dump retrieved from https://planet.osm.org. `https://www.openstreetmap.org`.

OpenStreetMap Wiki, 2024. Overpass api — openstreetmap wiki. URL: `https://wiki.openstreetmap.org/w/index.php?title=Overpass_API&oldid=2745759`. [Online; accessed 17-November-2024].

Statistics Canada, 2022. Census profile, 2021 census of population. `https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E`.

von Bergmann, J., Shkolnik, D., Jacobs, A., 2022. cancensus: R package to access, retrieve, and work with Canadian Census data and geography. URL: `https://mountainmath.github.io/cancensus/`. r package version 0.5.7.

Widener, M.J., Farber, S., Neutens, T., Horner, M., 2015. Spatiotemporal accessibility to supermarkets using public transit: an interaction potential approach in cincinnati, ohio. Journal of Transport Geography 42, 72–83.

Widener, M.J., Minaker, L., Farber, S., Allen, J., Vitali, B., Coleman, P.C., Cook, B., 2017. How do changes in the daily food and transportation environments affect grocery store accessibility? Applied geography 83, 46–62.

Yin, Z., 2024. geog712package: GEOG 712 R Package Activity. URL: `https://github.com/zehuiyin/geog712package`. r package version 0.1.1.