Using betweenness metrics to investigate the geographical distribution of retailers

Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
0(0) 1–18

DOI: 10.1177/2399808320971303
journals.sagepub.com/home/epb
SAGE

Using betweenness metrics to investigate the geographical distribution of retailers

Luigi Buzzacchi
Polytechnic of Turin, Italy

Philippe Leveque
Polytechnic of Milan, Italy

Roberta Taramino and Giulio Zotteri
Polytechnic of Turin, Italy

Abstract

In retailing, a location's accessibility and attractiveness depends on the spatial distribution of other stores and consumers. In particular, the literature shows that a place is more attractive for retailers if the generic routes taken by consumers often cross it. However, previous studies failed to consider that there are at least two possible consumer routes: job commutes from residential to workplaces and shopping trips among stores. In this paper, we analyze the impact of both consumer routes on the commercial patterns in Turin. The paper demonstrates that daily commutes to workplaces do not benefit a retailer along the trip, as much as journeys for shopping purposes do. In particular, we show that the benefits that a store can have when localized on the routes depend on the kind of goods it sells. Finally, the paper shows that stores selling homogeneous products and stores selling differentiated goods subject to comparison can differently benefit from being located in population hotspots and in commercial areas.

Keywords

Location, retail, network accessibility, betweenness

Corresponding author:
Luigi Buzzacchi, Interuniversity Department of Regional and Urban Studies and Planning, Polytechnic of Turin, FULL | The Future Urban Legacy Lab, Viale Pier Andrea Mattioli, 39, Torino 10125, Italy.
Email: luigi.buzzacchi@polito.it

Introduction

A considerable amount of literature has been published on retailing location; these studies assume that retailing plays a vital role in the metabolism of a city (Batty, 1997; Meltzer and Schuetz, 2012; Waldfogel, 2008); "brick and mortar" retailing is a locally-based economic activity, where both demand and supply contain spatial characterizations. Different strands of literature are focused on the analysis of the forces driving the distribution of retailers in the city (Stahl, 1987). In recent years, the competitive pressure of internet retailing has been transforming the competitive landscape and more and more consumers shop online, especially in urban areas (Farag et al., 2006), where the internet and delivery services are widely available. However, still a large fraction of total consumption happens in the brick and mortar stores (only 7.3% of total retail sales in 2019 occur online in Italy).[1] Still this ubiquitous competitive force puts pressure on weaker and less productive retailers that might exit the market, thus transforming the geographical distribution of stores in the cities.

Traditionally, the literature on retailing has suggested that stores should be located in areas with a high density of potential customers; empirical models (Seim, 2006; Waldfogel, 2008) have shown that stores tend to locate in areas with high density of people, i.e. easily accessible for consumers. Intensity of local consumption matters for all stores, but it plays a crucial role when consumers are unwilling to travel long distances in search for a product. This stream of research shows that demographics are a key driver of the attractiveness of any given urban area.

A second recent stream of literature investigates the impact of urban morphology and population mobility on store location; indeed, "every local market is different in its shape and its road network" and "these differences may have important impacts on the resulting market structure" (Aguirregabiria and Suzuki, 2016: 33).

First, stores can only be located in buildings that can actually accommodate them. Therefore, the distribution of buildings impacts store distribution. Moreover, in an urban context, people move and not all shopping trips start at the consumer's house. Thus, not only the spatial distribution of retailers and population but the impact of urban morphology and the flow of consumers within the city matter too (Oppewal and Holyoake, 2004; Teller and Elms, 2012).

Also, strategic interactions among firms determine location decisions; the economic literature has suggested that retailers might want to locate near other retailers to enjoy agglomeration economies (Stahl, 1987; Wolinsky, 1983); retailers might prefer locating near other retailers, because they jointly are more attractive to the consumers who do not know their exact tastes regarding commodities, who might want to compare products, prices, or find complementary items in a single shopping trip (Claycombe, 1991; Fischer and Harrington, 1996; Konishi, 2005). These agglomeration economies might draw more traffic to the set of neighboring stores (Koster and Pasidis, 2019). Agglomeration economies are counter-balanced by an increased level of competition among retailers, since consumers will switch from a retailer to the competitor more easily (i.e. with lower search cost). The first tend to group retailers together, the latter could tend to spread them evenly within the city (Baum and Haveman, 1997; Brown, 1993; Economides, 1986). Eaton and Lipsey (1979, 1982) and Fujita and Thisse (1996) provide a general treatment of equilibria, obtained from the strategic behaviors of consumers and firms.

The observed geographical distribution of retail stores depends on three processes: entry, location choice and exit. These processes are driven by the economic forces illustrated above as well as random events. Still, over time, the economic forces tend to overcome the random events and to shape the distribution of stores in the city. Indeed, retailers will be more inclined to locate/move where they expect to maximize their profits. Additionally, retailers who locate in such areas are more likely to survive than others.

In this paper, we investigate the data involving a static snapshot of retailers' distribution in the city of Turin and thus, we cannot disentangle the contribution of the entry, location, and exit processes. However, we can observe the distribution of stores and try to investigate the forces that contributed to it.

Namely, internet retailing is one of the forces that change the dynamics of the industry (see Weltevreden and Van Rietbergen, 2009). Clearly, the survival of brick and mortar stores might depend on the pressure from internet retailing in the product category they compete in, but all areas of the city will be subject to the same given competitive pressure from online retailers. Hence, when we observe the distribution of stores that were able to endure the competitive pressure of online retailers in any given category, we still can investigate their geographical distribution and wonder what spatial economic forces shaped the geographical distribution of stores. Weltevreden and Van Rietbergen (2009: 297) show that "accessibility and travel time" (i.e. geographical features of the brick and mortar stores) are a significant driver of substitution between brick and mortar shopping and online shopping in a European context.

The paper begins with a literature review, which provided us with some hypotheses to test. The subsequent section is concerned with the empirical tests. There, dependent and independent variables are defined and the available data are illustrated. Subsequently, we present our empirical strategy and the results obtained and the last section concludes the paper.

Research framework and hypotheses

The empirical literature focusing on the location of retailers within the city introduces several metrics for capturing location attributes that affect the probability of finding a retailer in a specific position (Guy, 1983; O'Kelly, 1999). Porta et al. (2009) and Sevtsuk (2014), in particular, use the *closeness*, *betweenness*, and *straightness* measures. While the closeness to customers and other stores can capture the demography and the agglomeration, straightness and betweenness can capture the impact of the urban morphology and the consumer flows within the city.

Conceptually, closeness indicates how far each location is from the others. When we compute the closeness of a location to consumers' houses, we can capture whether the location is conveniently located near places where people live (demographics). When we compute the closeness of a location to retail stores, we can measure the extent to which it might either enjoy economies of agglomeration or the presence of specific advantages of that location (e.g. sources of consumers' traffic such as large train stations).

The betweenness can capture whether people pass by a given location while moving within the city. Hence, the betweenness measure indicates the potential traffic passing by a given location and traffic is a key driver of the attractiveness of any given location for retailers.

Sevtsuk (2014) investigates the impact of the flow of people within the city on retail locations. The betweenness index used by Sevtsuk estimates the number of times a consumer passes by a location, using the shortest paths from every building to every building in the city. He demonstrates that such an index is a useful metric to investigate retail location distribution. He states that all retailers appear to value places with higher levels of passing traffic, i.e. high betweenness values.

Finally, the straightness metric assumes that shopping strategies often include unplanned purchases at more visible physical stores; places highly noticeable from the shopping path are consequently more frequented, making them more attractive as retailer locations.

In this paper, we analyze location attractiveness and accessibility, using all of the above metrics. Our contribution is twofold. First, we show that different flows of people can have different impacts on the location distribution of stores in a city. Second, we show that stores selling different kinds of goods might benefit from the different flows of consumers differently.

As for the first point, our research argues that a betweenness index where sources and destinations are all the buildings in a city, may not fully capture the type and the magnitude of consumer flows. Indeed, we argue that different kinds of trips might have different impacts on retail stores, thus, determining the real importance of "being in the middle" for the retail system.

On the one hand, we estimate the daily commutes of people from their home to their workspaces. These trips are relatively long and might be performed with faster means of transportation (e.g. cars, subways) and, thus, might not generate a large positive impact on consumer demands and attractiveness of the location for retailers.

On the other hand, we can simulate shopping trips from store to store; these trips are relatively shorter and typically on foot, thus making it easier for consumers to stop by and shop, increasing the attractiveness of the location for retailers.

Subsequently, for each building in the city, we will be able to calculate the betweenness for both the first (daily commutes from home to work) and second set of routes (shopping trips from store to store).

As for the second point, the literature about consumers' shopping behaviors highlights three possible conflicting habits (Carlson and McAfee, 1984; Schulz and Stahl, 1996; Urbany et al., 1996). Consumers want to minimize the time cost for searching for the goods they want to purchase. Also, they want to maximize the opportunity to compare goods, if they are purchasing nonhomogeneous products. Moreover, they try to combine purchases through "multipurpose shopping" using linked trips (Lambiri et al., 2017) that tends to increase the variety of complementary shops in a given area (Burdett and Malueg, 1981; McLafferty and Ghosh, 1986).

Also, part of the purchases of any given shopping trip are unplanned. They can also be decided in the store, driven by an in-store promotion or by a well-designed shopping window. These unplanned purchases make the high store locations (i.e. high consumer traffic areas) even more attractive.

Finally, both consumers and retailers appreciate to be located in dense commercial areas; the former wish to compare products and to combine purchases in linked trips, the latter benefit from the increased traffic of consumers.

The net balance between these forces driving consumers' behavior (and thus retailers' location choices) could depend on the type of goods that are bought (sold). Copeland, in 1923, outlined a basic distinction between convenience goods and shopping goods.[2] The first ones are essential consumer goods, characterized by lower product differentiations and lower price dispersions. The latter ones are differentiated according to some characteristics of the goods that the consumer typically compares before purchasing. This distinction between different types of goods is essential, since consumers tend to behave differently when shopping for convenience versus shopping goods. Thus, business retail strategies, among which the choice of location, must reflect the specific nature of the products sold (Holton, 1958).

The results of this literature suggest that:

- *Convenience goods* tend to be purchased from easily accessible stores (Bucklin, 1963). Indeed, consumers are expected to be relatively indifferent towards which store to visit, given their slight product differentiation. Consumers are typically familiar with these products and, as soon as they decide what convenience goods they want to buy, they will reduce the time cost spent searching. This is also because these products are, generally, relatively cheaper and purchased quite frequently. Hence, retailers who choose easily accessible locations for the consumers gain a sizable competitive advantage. These consumers will in fact tend to purchase such products from near their houses, their workplaces, or along the routes they frequent most.

- *Shopping goods* tend to be compared by consumers, who are pleased to visit many heterogeneous stores to compare products and prices before making their purchase. Consumers often compare products because they have not been able to establish exactly what they desire before the shopping trip and want to compare products and prices during the shopping trip (Bucklin, 1963). Consequently, the consumer appreciates a spatially concentrated distribution of shopping goods stores and retailers selling shopping goods can benefit from being near other such retailers, despite the increase in competition.

Therefore:

Hypothesis 1: The spatial distribution of the population has a positive impact on the spatial distribution of stores; the higher the number of people living close to a given location, the higher the probability of visiting a store in that given location.

Hypothesis 2: The density of retailers around a location is positively correlated to the attractiveness of that location for all retailers. This effect can be attributed first to the opportunity that the location offers to the consumers who want to purchase more than one good on a single trip to reduce search costs, second to the increased possibility to compare for consumers who want to compare before purchasing, third to the presence of specific advantages, such as infrastructures (Teller and Elms, 2012). Moreover, a store located in places where retailers agglomerate benefits a higher number of consumers visits and, hence, of unplanned purchases.

Hypothesis 3: The flow of people has a positive impact on the spatial distribution of stores; the higher the betweenness of a given location, the higher the probability of noticing a store in that given location. However, the traffic generated by shopping trips benefits the retail activities more than traffic generated by daily commutes, because daily commutes are performed on tighter schedules and with quicker means of transportation; hence, in case of daily commutes, stopovers and linked trips are less convenient. Moreover, the higher the time spent on shopping activity, the higher the possibility for stores to benefit from unplanned purchases that are decided in front of the store; we expect the impact of the betweenness of daily commutes, with respect to the betweenness of shopping trips on retail activity, to be lower.

Hypothesis 4: The visibility/accessibility of a specific location tends to make that location attractive for all retailers. In particular, a visible store can benefit from unplanned purchases.

Hypothesis 5: The density of retailers, the traffic generated by shopping trips, and accessibility have an impact that is higher for stores selling convenience goods than for stores selling shopping goods. Indeed, even if keeping other factors constant, almost all retailers prefer a location more accessible, with more shopping traffic passing by the front door and belonging to a shopping district, retailers who serve consumers who compare products tend to be planned destinations of shopping trips, while convenience goods stores benefit more from unplanned buying.

Data and dependent variables

Data

The analysis was carried out in the municipality of Turin, excluding the hilly area of the city. All variables are measured, if not differently stated, on 31 December 2016. The area studied corresponds to 109.4 km^2, with a population of 891,916 inhabitants.

Within this area, we obtained information about every active retail license (13,447);[3] among the others, their geographic coordinates and their merchandise category, which we categorized as convenience or shopping goods.[4]

Second, we obtained information on every one of the 37,394 buildings in the area.[5] We learned the geographic coordinates, the floor area and the volume of each building and their prevailing usage (residential, commercial, industrial, public administration, etc.). In particular, 28,026 out of the 37,394 buildings present exclusive or prevailing residential use.

The dependent variables we aim to investigate in all the analyses is the probability that a residential building hosts: (i) a retailer, (ii) a convenience retailer, or (iii) a shopping retailer; these three variables capture the distribution of retailers, given the distribution of buildings in the city of Turin.

Third, the geographic layer containing demographic information for each of the 3,439 census zones of Turin[6] provided the number of residents. Within each census zone, the population was attributed to each residential building, in proportion to the volume of the building. Hence, we obtained an estimate of the population living in each building.

Finally, we obtained information on the spatial distribution of travels between home and work in Turin, to capture the daily commutes of workers.[7]

Summing up, the database presents—for every building—information about retail presence, population, morphological attributes. Therefore, it offers a very detailed representation of the city's retail ecosystem.[8]

Demographic, behavioral, and morphologic metrics

In this section, we illustrate the variables that we associate to each location (building) for measuring its attractiveness as a retailer location. These variables combine all the morphological, demographic and behavioral information we have, as suggested by the theoretical hypotheses advanced in the previous section; we suppose that consumers trade off search costs and good/price comparisons when shopping, by purchasing either near their home, their workplace or along their typical routes.

Hence, below we present density indexes betweenness indexes and a straightness index. These indexes have been used for many purposes in urban design. Among the others, they have been applied to study their effect on residential dynamics (Omer, 2005) pedestrian volume (Kang, 2015) and to the public transportation network (Porta and Scheurer, 2006). In this paper, density indexes (density of the population and the retailers), betweenness indexes (betweenness of daily commutes and of shopping trips), and a straightness index will be measured for every location where we estimate the probability of finding a retailer, i.e. for every (prevailing or exclusive) residential building in Turin.

Population density index. The first variable that captures demography is the density of resident population, DP_i (density of population around building i)

$$DP_i(r) = \sum_{d_{ij} < r} P_j$$

where the sum is extended to P_j—the population of every building—provided that the distance d_{ij} between the reference building i and the building j is lower than the radius r. We will use this variable to test whether it is more likely for a retailer to locate where more people live.

Retailer density index. The second density variable adds agglomeration of stores to the story: it is the density of retailers around building i, DR_i

$$DR_i(r) = \sum_{d_{ij} < r} R_j$$

where the sum is extended to R_j—the stores located in every building j[9]—provided that the distance d_{ij} between the reference building i and the building j is lower than the radius r. We will use this variable to test whether retailers tend to flourish when located in clusters.

The DR variable incorporates at least two phenomena, other than agglomeration economies. On one hand, competition forces tend to push store owners to minimize or maximize the distance (Economides, 1986) from other competitors, as a function of the structural characters of the market where they operate;[10] on the other hand, density also measures the advantages of each location not captured by the other independent variables of our analyses. In this sense, DR_i can be considered as a control variable incorporating other various advantages of location i.[11]

Betweenness indexes. The third piece of the study deals with morphology and flow of people. People live not only at home but also in other places, such as workplaces. The betweenness variables introduce measures of the traffic of consumers along the routes of their daily travels within the city.

In general, a betweenness indexes B_i measures the traffic of people taking routes passing next to a specific position i (Porta et al., 2009; Sevtsuk, 2014). In general

$$B_i = \sum_j \sum_k \alpha^i_{jk} W_{jk}$$

examines all the routes from building j to building k; α^i_{jk} is equal to 1, if the path from j to k passes through i and 0 otherwise and W_{jk} is the (possibly estimated) number of individuals taking the j–k route. Our claim is that travelers behave differently, depending on the scope of their travel, so that different betweenness measures are expected to variously affect the probability of finding different types of retailers; in this sense, we are able to estimate the number of individuals along their travel for reaching their working location BW_i and along their shopping travels BS_i

$$BW_i = \sum_j \sum_k \alpha^i_{jk} W_{jk}$$

$$BS_i(r) = \sum_j \sum_{\substack{k \\ d_{jk} < r}} \alpha^i_{jk}$$

In our BW_i variable, W_{jk} is the estimate of the number of workers who live in building j and work in building k.[12] Similarly, in BS_i (r) we have included, with equal unitary weight, the shortest routes connecting each of j–k stores, located in different buildings, provided that the distance between j and k is smaller than r. In other words, we are attributing the same probability to each shopping trip between j and k, within a distance r, centered on building i. We will use these two variables to test whether retailers tend to flourish along routes covered by consumers when they go to work (BW_i) and shop (BS_i).

Straightness index. Finally, we also introduce a variable that measures the visibility of building i, a straightness measure STR_i. Indeed, a generic consumer shopping in the city is expected to be more attracted by stores that are visible along the shopping path, than by stores that are around the corner in parallel or perpendicular streets to the one that they are walking through. This idea can be captured by the following index proposed by Porta et al. (2009)

$$STR_i(r) = \frac{1}{N(r)} \sum_{\substack{j \\ d_{ij} < r}} \frac{d_{ij}}{D_{ij}}$$

where the sum is extended to every one of the $N(r)$ stores, whose (Euclidean) distance d_{ij} from building i is smaller than r. D_{ij} is the actual distance on the road network between building i and store j. STR_i, consequently, is higher when the stores close to the building under examination are located on a more *straight* connection with building i; so that we can say that the building is more visible from those stores. We will use this variable to test whether the store locations can be considered more attractive for retailers when more visible, in particular as target of unplanned purchases.

The two density indexes (population and retailers), the betweenness from store to store and the straightness index, are computed using predefined distances; we are aware that changing distances may change the impact of the index on the phenomenon we are investigating. However, the reasons why we opt for the 200 meter radius for the density of the population, 600 meters for the distance of shopping trips, and 100 meters for the straightness, are summarized below.[13]

On the one hand, the literature (see, e.g. Handy and Niemeier, 1997; Sevtsuk, 2014) suggests that typical consumers' shopping trips lay in areas corresponding to 10–15 minutes walks; therefore, we approximate both the relevant density of retailers and the betweenness from store to store as corresponding to distances of no more than 600 meters.

On the other hand, the density of population is introduced in the model to test whether consumers purchase, for some goods, from a store that is "near" their home, rather than going on a shopping trip. This type of purchases is normally performed within shorter time in "easily accessible" stores, corresponding to distances that we approximate to about 200 meters from their houses.

Finally, the straightness index aims at measuring the visibility of other stores from a particular building. One hundred meters is supposed to be a good approximation of the radius of action for the eyes of distracted consumers out-shopping.

The next section investigates the empirical relevance of the variables when trying to explain the distribution of stores in a city, using data from Turin. The results will reveal whether these metrics provide insights regarding the spatial distribution of stores in an urban context.

Methodology and results

This research aims at estimating the probability of a building to host a retailer, a retailer selling convenience goods, or a retailer selling shopping goods, given the set of building characteristics (demographics, mobility behaviors and spatial form). In Table 1, the summary statistics of dependent (a) and independent (b) variables are provided.[14] The test of hypotheses 1–5 is carried out using probit regression models.

Table 2a shows the outcome of a probit model, where the dependent variable is a dummy variable, which equals 1 if the building hosts at least one retailer. In Table 2b, the dependent variable is a dummy variable, which equals 1 if the building hosts at least one convenience goods retailer. Finally, in Table 2c, the dependent variable is a dummy variable, which equals 1 if the building hosts at least one shopping goods retailer.

Before we comment Table 2 to test hypotheses 1–5, the robustness and significance of the models deserve a comment. All models are very significant, according to both likelihood and the Wald tests. Furthermore, when we compare the coefficients of Models 1, 2, and 3, in every single table, they are (in most cases) very significant. This is due to the large amount of observations in the database ($n = 28{,}026$). Moreover, coefficients are stable across Models 1, 2, and 3, but they are different across the probit models in Table 2 according to the different dependent variables. Thus, the models are statistically significant and stable. Moreover, the coefficients of some variables significantly differ across tables; the divisions we made in terms of retailers, retailers selling convenience goods and retailers selling shopping goods, generate different results and insights and thus proves to be a relevant one.

In the rest of the section, the results will be described according to the list of hypotheses from "Research framework and hypotheses" section. When not differently specified, we will refer to Models 2. The coefficients of the models are not elasticities, hence, marginal effects at the mean (*dydx* measures in the tables) will also be discussed. Moreover, along with the discussion, we will also provide (in relevant cases) the partial effects of the discrete variables (50th percentile–95th percentile).

Hypothesis 1: Table 2a confirms the hypothesis that population density supports retail activity; the surrounding population has a positive impact on the probability that a building hosts one or more retailers. Indeed, the impact is positive and significant in all three models. The result holds for convenience retailers as well (Table 2b) and the impact of the resident population is even larger than for the average retailer. As an additional measure of the effect, an average building (i.e. here and below, a building with median values of the independent variables) but characterized by high values of *DP* (95th percentile), has a 0.9% higher chance of hosting at least one retailer and a 2.0% higher chance of hosting a retailer selling convenience goods. This is consistent with the fact that for convenience goods, consumers should pay more attention to the minimization of travel costs in their shopping choices.

Interestingly, the impact of population on the probability of finding a shopping goods retailer in a building is significantly negative (Table 2c). Indeed, to reduce search and comparison costs to consumers, shopping goods retailers tend to concentrate in accessible zones, which tend to prevail in areas with relatively low population density.

Hypothesis 2: All the models confirm the hypothesis that retail density supports retail activity: the coefficients of *DR* are positive and very significant and stable in all three models. Many retail models focus on the distribution of the population (i.e. where people live, as in Hp. 1); our evidence shows that the distribution of retail stores matters more than the distribution of residents in the attractiveness of locations for retailers. *Model 3* in Table 2a shows that ceteris paribus, the effect of an increase for an average building at the 95th percentile in the *DP* and *DR* variables is an increase of 0.9% and 14.6% in the probability to host a store, respectively.

This evidence might be due to strong agglomeration effects; they seem to prevail over possible competition forces that lead retailers to locate far from each other. Moreover, consider that the wide set of product categories in the database is such that competition

Table 1. Summary statistics.

(a): Dependent variables	Yes	No	Total
Presence of one or more retailers in the building	5670	22,356	28,026
Presence of one or more convenience goods retailers in the building	4624	23,402	28,026
Presence of one or more shopping goods retailers in the building	2429	25,597	28,026

(b): Independent variables	Label	Min	Max	Mean	5th percentile	Median	95th percentile	Standard deviation
Density of population within a 200 meters radius	*DP*	2.0	5390.0	2177.8	499.7	2204.9	3953.5	1057.2
Density of retailers within a 600 meters radius	*DR*	1.0	1257.0	261.7	41.0	234.0	570.0	188.3
Betweenness of daily commutes[a]	*BW*	0.0	12.7	0.2	0.0	0.0	0.4	0.7
Betweenness of shopping trips within a 600 meters radius[b]	*BS*	0.0	12.9	0.6	0.0	0.2	2.4	0.9
Straightness within a 100 meters radius	*STR*	0.00	1.00	0.67	0.42	0.70	0.84	0.14

[a]The *BW* index is normalized to the total number of people getting to workplaces (so that it is the share of workers of Turin passing next to the building) and, then, it is multiplied by 1000.

[b]The *BS* index is normalized to $[(N-1)(N-2)]$ (so that it is the probability that a generic shopping route from each couple of stores of the city will pass next to the building, considering that only the routes within a radius of 600 meters will be actually covered) and, then, it is multiplied by 100,000.

Table 2. The probability for a building to host a retailer. Probit models ($n = 28{,}026$).

	Model 1		Model 2		Model 3	
	Coeff.	dydx	Coeff.	dydx	Coeff.	dydx
(a): All retailers						
Constant	−1.460 (0.023)***		−1.540 (0.023)***		−1.805 (0.053)***	
DP	0.343***	0.09	0.203 (0.094)*	0.05	0.206 (0.094)*	0.05
DR	0.192 (0.005)***	0.05	0.159 (0.005)***	0.04	0.159 (0.005)***	0.04
BW			−0.110 (0.014)***	−0.03	−0.105 (0.014)***	−0.03
BS			0.309 (0.010)***	0.08	0.304 (0.010)***	0.08
STR					0.398 (0.071)***	0.11
Pseudo R^2	0.073		0.109		0.110	
Log-likel.	−13,086.57		−12,578.38		−12,562.31	
LR (χ^2)	2054.14***		3070.52***		3102.67***	
Wald (χ^2)	1919.85***		2794.55***		2819.67***	
Correct pred. (%)	74.10		76.68		76.69	
(b): Convenience goods retailers						
Constant	−1.583 (0.024)***		−1.661 (0.025)***		−1.967 (0.056)***	
DP	0.645 (0.094)***	0.15	0.520 (0.097)***	0.12	0.523 (0.097)***	0.12
DR	0.160 (0.005)***	0.04	0.127 (0.005)***	0.03	0.126 (0.005)***	0.03
BW			−0.109 (0.014)***	−0.02	−0.104 (0.014)***	−0.02
BS			0.291 (0.010)***	0.07	0.287 (0.010)***	0.07
STR					0.459 (0.075)***	0.10
Pseudo R^2	0.060		0.095		0.097	
Log-likel.	−12,578.38		−11,358.50		−11,339.47	
LR (χ^2)	1503.54***		2385.99***		2424.05***	
Wald (χ^2)	1450.16***		2251.29***		2280.14***	
Correct pred. (%)	77.16		79.48		79.50	
(c): Shopping goods retailers						
Constant	−1.904 (0.030)***		−1.957 (0.030)***		−2.075 (0.067)***	
DP	−0.168 (0.112)	−0.02	−0.267 (0.115)*	−0.04	−0.268 (0.115)*	−0.04
DR	0.189 (0.005)***	0.03	0.166 (0.005)***	0.02	0.166 (0.005)***	0.02
BW			−0.069 (0.017)***	−0.01	−0.067 (0.017)***	−0.01
BS			0.202 (0.011)***	0.03	0.200 (0.011)***	0.03
STR					0.178 (0.090)*	0.02
Pseudo R^2	0.086		0.106		0.106	
Log-likel.	−7553.32		−7386.60		−7384.62	
LR (χ^2)	1415.46***		1748.90***		1752.86***	
Wald (χ^2)	1386.10***		1695.02***		1698.07***	
Correct pred. (%)	87.41		87.89		87.90	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

occurs within small subsets of competing retailers (e.g. competition among hardware stores[15]) and thus, shall be rather weak in our framework. On the other hand, stores selling complementary merchandise categories or merchandise categories subject to shopping comparison might be generating proximity externalities. This makes agglomeration forces rather strong. However, we already mentioned that this outcome could also be consistent with the concentration of retailers in locations with some advantage (e.g. railway stations generating large traffic of daily commuters.

Hypothesis 3: The evidence provided shows that the betweenness of daily commutes and the betweenness of shopping trips have an opposite effect. In every model, the coefficients of BW are negative and significant, while the coefficients of BS are positive and significant. This confirms the basic hypothesis behind this paper: while previous literature (Sevtsuk, 2014) has shown that the city form leads people flow and these have an impact on the distribution of retail activity, this paper shows that the contribution of such flows shall be disentangled since their nature produces very different impacts. From Table 2a, on the one hand, an average building characterized by high values of BW has a 1.0% lower chance of hosting a retailer. On the other hand, the average building characterized by high values of BS is 21.1% more likely to host a retailer. Table 2b and Table 2c show similar results for convenience and shopping goods. Notice that the impact of the population is also lower than the impact of the ease of access, as measured by the betweenness of shopping trips. In other words (at the least at the urban scale), the distance from where people live seems to matter less than the patterns of trips within the city and the ease with which the location can be accessed.

Hypothesis 4: In all the cases, Models 3 show that the straightness, i.e. the visibility of stores, has a positive and statistically significant effect on the likelihood to find at least one retailer in a given residential building.

In conclusion, the shape of the urban form (straightness), the dynamic patterns of consumer movements within the city (betweenness of shopping trips), and the spatial distribution of retailers matter even more than a static perspective on where potential consumers live.

Hypothesis 5: The results we have illustrated so far have in general the same sign for convenience and shopping goods (see Table 2b and 2c). The only exception is in the specific case of the density of population, which has a positive and large impact in the case of convenience goods, while in the case of shopping goods, the impact is small and negative. An average building that has the 95th percentile value of DP is 2.0% more likely and 0.6% less likely to host a convenience goods store and a shopping goods one, respectively.

In general, the impact of all the variables proves to be much more relevant in magnitude for convenience stores, than for stores selling shopping goods.

In particular, convenience goods stores benefit much more than shopping goods stores from higher values of betweenness of shopping trips and straightness: an average building that has the 95th percentile value of BS and STR, is, respectively 17.8% and 1.7% more likely to host a convenience goods store, while these figures are reduced to 7.5% and 0.4% for shopping goods stores. We interpret this evidence with the special role of unplanned purchases among convenience goods.

Concluding remarks

While writing these conclusive notes (April 2020), most of us are confined to work in our homes and we can only visit them to purchase groceries and essential items in the area near our homes. The Covid-19 pandemic has imposed significant, and differentiated, changes to our ways of living, and in particular to households spending patterns and movements. Several interesting academic papers already appear, offering first analyses—in the Covid-19 regime—of demographic differences in mobility (Coven and Gupta, 2020), of the ability to survive of small businesses (Bartik et al., 2020) and of the new pattern of consumption and of the shift from brick and mortar to online retailers (Baker et al., 2020). Of course, we do not know whether the situation will return to the "old" regime and when.

Our paper, however, is well suited to understand the possible effects of modified consumption behavior in the metabolism of the city since it investigates in detail not only the role of the density of residents and stores, but also the flow of commuters and the flow of consumers from store to store for driving competition and location of retailers.

Previous researches showed that stores tend to concentrate where people live (e.g. Seim, 2006; Waldfogel, 2008), while Sevtsuk (2014) proved that, not only where people live but also the flow of people in the city network tends to drive the spatial distribution of retailers.

Thus, often planning tends to investigate ratios between residents and stores, and our paper suggests that the internal dynamics of the retail industry and the flow of people, rather than just the static distribution of where people live, shall be considered when planning retail spaces. Moreover, these findings might be of interest for retailers. Indeed, the results demonstrate the features that make a location attractive and provide the metrics to measure them.

Moreover, the paper shows that there are different types of flows of people in the city and their contribution shall be disentangled. Indeed, while the contribution of the daily commutes in the city is negative, the contribution of shop to shop trips is positive, significant and large. Again, this fundamental understanding of what helps retail activities thrive can contribute not only to the knowledge, but also to the planning and location choices of retail investors.

In our paper, these metrics on the flow of people in the city are proxies of actual trips, estimated from publicly available data. Though the significance of the metrics and estimates are proved by the robustness of the results, the relevance of the results deserve further investigation and the use of actual traffic data—collected from cell-phone GPS records—which might provide additional opportunities for further research. Such data may pave the way for various, more detailed analyses, such as the impacts of hourly people mobility, new large store openings and other sources of traffic (e.g. underground stations).

This paper also investigates the impact of the urban form and proves that straightness matters; in other words, the visibility of one store from other stores increases the probability of a building hosting a store.

Finally, this paper also shows that the above phenomena vary for different kinds of stores. Namely, stores selling convenience goods differ from stores selling shopping goods. Interestingly, the density of population increases the probability of a building hosting a store (as suggested by literature). However, the density of population lowers the probability of a building hosting a store selling shopping goods. Indeed, these latter stores are destinations of ad hoc shopping trips. This result is consistent with another observation: stores selling shopping goods prevail in shopping malls (that are often located in suburban areas and are not included in our database) with high concentration of stores that tend to be located outside of cities, in areas where the population density is relatively low. Low population density is not a desirable attribute *per se* for any retail location. However, given that these retailers simply cannot survive out of local shoppers, accessibility matters much more than local density of population (Ertekin et al., 2008; Han, 2014) and thus these retail activities tend to be located in areas with great accessibility (both city centers through the road network and public transportation and malls through highways) at the expenses of low population density.

This result has potential implications for urban planners. Indeed, it seems to suggest that large concentrations of stores selling shopping goods can be created, regardless of the concentration of residents-consumers, since consumers are willing to travel both to suburban malls and concentrations of stores in the city, to the extent that these concentrations are easily accessible.

Furthermore, the paper raises a few more research questions.

First, in all the spatial analyses, the scale of the analyses matter. This paper shows that the impact of the distribution of population is weaker than other factors (density of stores, betweenness, and straightness) at the urban scale, but one wonders whether the same conclusion holds for longer distances. Thus, one might want to test the same hypotheses at the province rather than at the city scale. In that case, the impact of the distribution of the population might be more compelling.

Second, in our analyses, we grouped the stores in two large sets: convenience stores and stores selling shopping goods. Given the breadth of our analyses, the impact of agglomeration (e.g. due to complementarity of stores) and natural advantage outweigh the impact of competition. The density of stores tends to have a positive impact on the probability that a building will host a store. One wonders if the same results hold when we consider smaller (though more consistent) samples of stores selling similar and possibly, substitute products. In this case, the number of observations drops dramatically and the statistical significance of results might be an issue.

Third, one wonders whether the results, though very robust empirically, partially depend on the sample. Thus, the comparison of data from Turin with data from other cities might provide further interesting insights.

Fourth, one might take a more dynamic approach and investigate the changes in the spatial distribution of stores with studies on entries/exits. One could also correlate such changes with the evolution of online shopping and check whether the products categories that are bought more often online are subject to more changes in the geographical distribution of stores.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We gratefully acknowledge financial support from the FULL – Future Urbal Legacy Lab, Politecnico di Torino.

ORCID iD

Roberta Taramino https://orcid.org/0000-0001-8592-1287

Notes

1. Source: *Osservatorio E-Commerce B2C, Politecnico di Milano*.
2. From Copeland (1923) on, goods are typically classified in to three categories: convenience goods, shopping goods, and luxury goods. Luxury goods are characterized by a strong brand reputation, recognized quality, and a price premium. The behavior of consumer when purchasing shopping or luxury goods is not very different, in our perspective: high willingness to compare and lower cost-opportunity for the time devoted to shopping. Not surprisingly, our empirical tests discussed in the results section indicate that luxury and shopping retailers' behavior is not statistically different from a location point of view. Hence, we decided to associate luxury goods to dhopping goods. Empirical tests that distinguish luxury from shopping are available upon request.

3. A licence is an open trade permit to sell a specific merchandise category related to a geographical address and a store size. The licence data set can be downloaded from http://geoportale.comune.torino.it/web/.
4. All the stores have been classified either as shopping goods stores or convenience goods stores. *Shopping goods*: clothing, households goods, sportswear, jewellery, footwear, furnishing accessories, lingerie, numismatics and philately, precious objects, artworks, glasses, fabrics, second-hand goods. *Convenience goods*: grocery, biological grocery, funeral articles, pet supplies, sanitary wear, audiovisual products, soft drinks, fuel, stationery, personal- and home-care products, paint shop, electrical households appliances, electronic items, wine, herbalist, drugstore, hardware store, flowers, photography, fruit and vegetables, ice-creams, toys, newspapers, informatics, hypermarket, bookstore, butchery, minimarket, fish market, phone center, pizza, perfumery, car parts and accessories, sexy shop, supermarket, tobacco shop, telephony, bakery, pastry.
5. Source: *Regional Technical Maps* of the *Laboratory of Analysis and Urban Territorial Representations* (http://www.lartu.polito.it/).
6. Source: Turin Geoportal (http://geoportale.comune.torino.it/web/).
7. *5T*, the Turin Transport Agency, occasionally produces an origin–destination matrix for traffic flows of workers in Turin. Based on interviews to residents and data from traffic detectors, the matrix measures the flows of individuals of residents who are getting to work between 166 areas of the city (these areas thus correspond to aggregations of (on average) 20 census zones). Consequently, we can estimate the flows of residents from area j to their work location in area k (j, $k = 1 \ldots 166$). The origins of the trips are the buildings of the area k in proportion to the residents of each building, while destinations are randomly attributed to the buildings in area k (in proportion to their volume).
8. Data descriptives are available upon request.
9. The sum here is extended to every building in the city, not only to the residential buildings represented in our dependent variable. In other words, DR_i correctly considers all of the retail licences active in the city since even the flow of consumers coming from stores located in commercial buildings can influence the probability that a residential building hosts a store.
10. Competition forces affect location decisions of retailers belonging to the same sector (e.g. pharmacy locations are affected by the distribution of other pharmacies). As we observe retailers' locations of large sets of retailers, in each set, competition forces are expected to impact the distribution we observe mildly. Indeed, the vast majority of retailers in each set are not competing directly.
11. The "retailers" considered in the definition of the DSi variable (and in the definition of the shop-to-shop betweenness variable BSi below) include only "proper" retailers belonging to the NACE sector 47.1-7 (*Retail trade, except of motor vehicles and motorcycles*), while excluding *Food and beverage service activities* (NACE sector 56), *Repair of computers and personal and household goods* (NACE sector 95), and *Other personal service activities* (NACE sector 96). The attractivity of a specific location generated by the surrounding commercial activity of course also benefits from traffic flows determined by all the types of stores, including those offering the services listed above. However, our variables are a very good proxy of the commercial liveliness (shopping traffic) of a specific location given that retailers are the majority of the commercial activities in Turin (about 55% of the licenses) and they are highly co-agglomerated with the other commercial activities.
12. Thanks to the *5T* origin–destination matrix and our estimate of the number of residents in each building (see again footnote 7) we have a building-to-building matrix that contains the number of individuals daily commuting for work between each couple of buildings in the city. The routes of every home-to-work travel have been obtained using a "shortest path algorithm"; we were then able to calculate our betweenness measure BWi, i.e. the normalized number of individual trips passing next to building i.
13. We have also conducted sensitivity analysis using alternative radii around 200 meters (density of population), 600 meters (distance of shopping trips), and 100 meters (straightness) and results are not significantly changing. Sensitivity checks are available upon request.
14. The correlation matrix is available upon request.

15. Competition strategies of firms in single product categories when they decide to locate in the geographic space could be actually very complicated (Huang and Levinson, 2011). Baum and Haveman (1997), for example, show that location decisions in the Manhattan hotel industry determine a distribution where hotels of similar price tend to agglomerate in order to avoid the hazards of localized price competition, while competition pushes hotels far apart from hotels of similar size.

References

Aguirregabiria V and Suzuki J (2016) Empirical games of market entry and spatial competition in retail industries. In Basker E (ed.) *Handbook on the Economics of Retailing and Distribution*. Cheltenham: Edward Elgar Publishing, pp.201–232.

Baker SR, Farrokhnia RA Meyer S, Pagel M, et al. (2020) How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic.National Bureau of Economic Research working paper series, n. w26949.

Bartik A, Bertrand M, Cullen Z, et al. (2020) How are small businesses adjusting to COVID-19? Early evidence from a survey. National Bureau of Economic Research working paper series, n. w26989.

Batty M (1997) The retail revolution. *Environment and Planning B: Planning and Design* 24(1): 1–2.

Baum JAC and Haveman HA (1997) Love thy neighbor? Differentiation and agglomeration in the Manhattan hotel industry, 1898-1990. *Administrative Science Quarterly* 42(2): 304–338.

Brown S (1993) Retail location theory: The legacy of Harold Hotelling. *International Review of Retail, Distribution and Consumer Research* 3(2): 185–229.

Bucklin LP (1963) *Classification of consumer goods*. *Journal of Marketing* 27(1): 50–55.

Burdett K and Malueg DA (1981) The theory of search for several goods. *Journal of Economic Theory* 24(3): 362–376.

Carlson JA and McAfee RP (1984) Joint search for several goods. *Journal of Economic Theory* 91(3): 337–345.

Claycombe RJ (1991) Spatial retail markets. *International Journal of Industrial Organization* 9(2): 303–312.

Copeland MT (1923) Relation of consumers' buying habits to marketing methods. *Harvard Business Review* 1(3): 282–289.

Coven J and Gupta A (2020) Disparities in mobility responses to covid-19. *NYU Stern Working Paper*.

Eaton BC and Lipsey RG (1979) Comparison shopping and the clustering of homogeneous firms. *Journal of Regional Science* 19(4): 421–435.

Eaton BC and Lipsey RG (1982) An economic theory of Central places. *The Economic Journal* 92(365): 56–72.

Economides N (1986) Minimal and maximal product differentiation in Hotelling's. *Economics Letters* 21(1): 67–71.

Ertekin O, Dokmeci V, Unlukara T, et al. (2008) Spatial distribution of shopping malls and analysis of their trade areas in Istanbul. *European Planning Studies* 16(1): 143–156.

Farag S, Weltevreden J, Van Rietbergen T, et al. (2006) E-shopping in The Netherlands: Does geography matter? *Environment and Planning B: Planning and Design* 33(1): 59–74.

Fischer JH and Harrington J Jr (1996) Product variety and firm agglomeration. *The RAND Journal of Economics* 27(2): 281–309.

Fujita M and Thisse JF (1996) Economics of agglomeration. *Journal of the Japanese and International Economies* 10(4): 339–378.

Guy CM (1983) The assessment of access to local shopping opportunities: A comparison of accessibility measures. *Environment and Planning B: Planning and Design* 10(2): 219–237.

Han D (2014) *The location pattern of malls in Texas*. PhD Thesis, Texas A&M University, USA.

Handy SL and Niemeier DA (1997) Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A: Economy and Space* 29(7): 1175–1194.

Holton RH (1958) The distinction between convenience goods, shopping goods and specialty goods. *Journal of Marketing* 23(1): 53–56.
Huang A and Levinson D (2011) Why retailers cluster: An agent model of location choice on supply chains. *Environment and Planning B: Planning and Design* 38(1): 82–94.
Kang CD (2015) The effects of spatial accessibility and centrality to land use on walking in Seoul, Korea. *Cities* 46: 94–103.
Konishi H (2005) Concentration of competing retail stores. *Journal of Urban Economics* 58(3): 488–512.
Koster H and Pasidis IN (2019) Shopping externalities and retail concentration: Evidence from Dutch shopping streets. *Journal of Urban Economics* 114: 103194.
Lambiri D, Faggian A and Wrigley N (2017) Linked-trip effects of 'town-Centre-first' era foodstore development: An assessment using difference-in-differences. *Environment and Planning B: Urban Analytics and City Science* 44(1): 160–179.
McLafferty SL and Ghosh A (1986) Multipurpose shopping and the location of retail firms. *Geographical Analysis* 18(3): 215–226.
Meltzer R and Schuetz J (2012) Bodegas or Bagel shops? Neighborhood differences in retail and household services. *Economic Development Quarterly* 26(1): 73–94.
O'Kelly ME (1999) Trade-area models and choice-based samples. *Methods. Environment and Planning A: Economy and Space* 31(4): 613–627.
Omer I (2005) How ethnicity influences residential distributions: an agent-based simulation. *Environment and Planning B: Planning and Design* 32(5): 657–672.
Oppewal H and Holyoake B (2004) Bundling and retail agglomeration effects on shopping behavior. *Journal of Retailing and Consumer Services* 11(2): 61–74.
Porta S, Strano E, Iacoviello V, et al. (2009) Street centrality and densities of retail and services in Bologna, *Italy. Environment and Planning B: Planning and Design* 36(3): 450–465.
Porta S and Scheurer J (2006) Centrality and connectivity in public transport networks and their significance for transport sustainability in cities. *World Planning Schools Congress* 13–16.
Schulz N and Stahl K (1996) Do consumers search for the highest price? Oligopoly equilibrium and monopoly optimum in differentiated-products markets. *The RAND Journal of Economics* 27(3): 542–562.
Seim K (2006) An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics* 37(3): 619–640.
Sevtsuk A (2014) Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research* 34(4): 374–392.
Stahl K (1987) Therories of urban business location. In: Mills ES (ed.) Handbook of Regional and Urban Economics. New York: Elsevier BV, pp.759–820.
Teller C and Elms JR (2012) Urban place marketing and retail agglomeration customers. *Journal of Marketing Management* 28(5–6): 546–567.
Urbany JE, Dickson PR and Kalapurakal R (1996) Price search in the retail grocery market. *Journal of Marketing* 60(2): 91–104.
Waldfogel J (2008) The median voter and the median consumer: Local private goods and population composition. *Journal of Urban Economics* 63(2): 567–582.
Weltevreden WJ and van Rietbergen T (2009) The implications of e-shopping for in-store shopping at various shopping locations in The Netherlands. *Environment and Planning B: Planning and Design* 36(2): 279–299.
Wolinsky A (1983) Retail trade concentration due to consumers' imperfect information. *The Bell Journal of Economics* 14(1): 275–282.

Luigi Buzzacchi is full Professor at Polytechnic of Turin (Italy), where he teaches courses in Urban Economics and Industrial Organization. His research interests lie in the area of

urban and regional economics, insurance and financial markets, firm size ad spatial competition.

Philippe Leveque born in Aosta (Italy) in 1993, is a PhD fellow of Polytechnic of Milan (Italy). He started his journey within academy in 2018, and his research project deals with Innovation and Policy for the Circular Use of Water Resources. Philippe got his master degree in Management Engineering with a thesis on the drivers of retailers' location choices.

Roberta Taramino earned her undergraduate degree in Engineering and Management at Polytechnic University of Torino, Italy. Roberta is actually a PhD candidate at the Polytechnic of Turin (Italy) and she is a teaching assistant in Marketing and Urban and Regional Economics. Roberta's research interest focuses on the relationship between retail location and urban morphology.

Giulio Zotteri Giulio is full Professor at Politecnico di Torino where he currently teaches Distribution Logistics, Marketing, and Urban and Regional Economics. Giulio has published more than 20 articles in international journals on the above topics. Giulio is part of the Consortium on Operational Excellence in Retail and has studied the retail sector for more than 20 years.