

Do Agglomeration Externalities Form Service Clusters? Evidence from a Location Lottery*

Yuki Higuchi¹, Ngan Bich Le², and Mari Tanaka^{3,2}

¹Sophia University

²Hitotsubashi University

³University of Tokyo

December 15, 2025

Abstract

Do agglomeration externalities drive the formation of service clusters, even when firms compete in close proximity? This paper provides causal evidence from one of the largest wholesale apparel and fabric markets in Hanoi. After a fire destroyed the market building in 1994, slots in the reconstructed building were allocated by lottery among previous owners, generating an essentially random initial distribution of products. Despite this randomness, product clusters formed rapidly and have continued to agglomerate over two decades. Using detailed shop-level survey data, we show that initial coincidental agglomeration strongly predicts the subsequent adoption of similar products in neighboring stores. Clustering arose through both within-owner product switching and ownership transfers. Spillovers of products across distant stores within the same owner are limited, suggesting that agglomeration externalities are more likely to be driven by demand-side factors (e.g., shopping externalities) than by supply-side factors (e.g., knowledge spillovers). These findings highlight the strength of agglomeration externalities in shaping service cluster formation, even in the presence of local competition and adjustment frictions.

Keywords: agglomeration externalities, service clusters, Vietnam, location lottery

*We thank Marco Gonzalez-Navarro, Kensuke Teshima, Kentaro Nakajima, Junichi Yamasaki, and the seminar participants in various academic workshops and conferences for their valuable comments and suggestions. We gratefully acknowledge financial support from JSPS Grant-in-Aid for Scientific Research (No. 24K00257). Finally, we sincerely thank the survey enumerators, Vu Hoang Nam, Dong Xuan Market Corporation, and survey respondents for their valuable inputs to this study. All errors are our own.

1 Introduction

Clusters of retail and service firms are a defining feature of large cities around the world. From the fashion districts of New York, Los Angeles, and Paris to the electronics streets of Chennai, Mumbai, Hong Kong, and Tokyo, similar businesses frequently cluster side by side in narrow geographic areas.¹ Their coexistence raises a long-standing question in urban and industrial economics: why do competitors choose to operate in such close proximity, despite intensified local competition? A large theoretical literature, dating back to Marshall (1920), highlights the role of *agglomeration externalities*—product-market and knowledge spillovers that make firms more productive or profitable when located near others in the same industry. On the demand side, co-location can attract more consumers by offering greater product variety or lower search costs. On the supply side, proximity may foster information exchange about industry-specific knowledge. These ideas are supported by numerous empirical studies documenting externalities within the service sector (e.g., Shoag and Veuger (2018), Nakajima and Teshima (2020), and Atkin et al. (2022)).

However, the existence of externalities does not necessarily imply that firms will endogenously form clusters. Firms must weigh the private benefits from proximity against countervailing forces: intensified price competition and adjustment costs such as changing products or acquiring new premises (e.g. land lot). Whether agglomeration externalities form service clusters, therefore, remains an empirical question. This issue is also highly relevant to urban-development policy. Many municipal governments attempt to promote or restrict retail concentrations through zoning and land-use regulation. If clustering arises spontaneously through market forces, such interventions may be unnecessary or even distortive.

Empirically identifying the causal effect of agglomeration externalities on cluster formation is challenging because firm location choices are typically influenced by unobserved local amenities, infrastructure, or customer and supplier locations. We address this challenge by exploiting a natural experiment in a large wholesale market building in Hanoi (known as Dong Xuan Market) that hosts more than 1,000 apparel and fabric wholesalers. Following a major fire in 1994, the building was reconstructed and reopened in 1997, and the store slots in the new building were allocated by lottery among previous shop owners. This lottery generated an effectively random initial spatial distribution of product types, enabling us to examine whether and how clusters emerged and evolved from a random starting point.

We document that despite random initial placement, clusters of similar products emerged

¹These examples refer to apparel/fashion districts in New York and Los Angeles, and Le Marais in Paris, and electronic streets in Ritchie Street in Chennai, Lamington Road in Mumbai, Sham Shui Po area in Hong Kong, and Akihabara town in Tokyo.

quickly and have intensified and persisted for almost twenty years. Using slot-level panel data constructed through our retrospective survey to shop owners, we first confirm that the initial slot allocation by the lottery appears to be random. We then show that neighborhoods that happened to contain a higher pre-lottery concentration of a product category (e.g. men’s T shirts) were subsequently more likely to see nearby shops sell the same category. These effects appear within one year of the market’s reopening and strengthen over time. This evidence suggests that externalities across stores formed clusters.

Investigating the mechanisms behind product changes, we find that both product switching within existing store owners and changes in store ownership played important roles. In particular, immediate responses occurred through rapid product switching by original owners, while more gradual adjustments emerged through subsequent ownership transfers.

We also examine the mechanisms through which agglomeration externalities operate, distinguishing between demand-side forces (such as shopping externalities that attract customers) and supply-side forces (such as knowledge spillovers across firms). Our evidence indicates that spillovers of product choices across distant stores owned by the same firm are limited, suggesting that knowledge spillovers are unlikely to be the primary driver of clustering in our setting.

Our study is linked to two strands of literature. First, our study is closely related to the empirical studies providing evidence suggesting the presence of agglomeration externalities in service sector (Arzaghi and Henderson, 2008; Combes et al., 2012; Shoag and Veuger, 2018; Benmelech et al., 2019; Bernstein et al., 2019; Koster et al., 2019; Nakajima and Teshima, 2020; Atkin et al., 2022; Baum-Snow et al., 2024). Among these, the study most closely related to ours is Nakajima and Teshima (2020). They exploit a lottery-based allocation of slots in the Tokyo Tsukiji fish market and show that greater diversity of neighboring firms improves the performance of small and specialized sellers. Because lotteries in the Tokyo fish market occur every 4–10 years, firms expect future reallocations and thus have limited incentives to make long-term adjustments such as changing product lines and transferring slot ownership. By contrast, the market we study in Hanoi held a one-time lottery nearly twenty years ago. This setting provides an ideal environment for studying the formation of clusters through agglomeration externalities.

Second, our main contribution is to the empirical literature on how agglomeration externalities affect formation of clusters or firm location choices (Ellison et al., 2010; Leonardi and Moretti, 2023; Baum-Snow et al., 2024; Vitali, 2024). While this has been a central question in urban economics, the causal evidence has been scarce. For instance, Leonardi and Moretti (2023) exploit abolishment of minimum distance requirement for restaurants in Milan and show that restaurants started to agglomerate in some neighborhoods after the deregulation.

Vitali (2024) examines the role of consumer information frictions on firms’ location decisions by estimating structural model in the setting of garment sector in Uganda. This study contributes to this literature by providing evidence on how agglomeration externalities affect formation of clusters based on a lottery-based natural experimental setting.

The remainder of the paper is organized as follows. Section 2 describes the institutional background of the wholesale market building in Hanoi. Section 3 introduces the survey data. Section 4 outlines the empirical strategy and demonstrates the randomness of the lottery-based allocation. Section 5 presents the main results, and Section 6 examines the mechanisms underlying product adjustments and the sources of agglomeration externalities. Section 7 concludes with policy implications and directions for future research.

2 Institutional Background

2.1 Overview of Dong Xuan Market

Dong Xuan Market is Hanoi’s one of the largest wholesale markets for apparel and fabrics. The three-story building accommodates more than one thousand wholesale shops, each occupying a small slot averaging a few square meters. Each floor covers roughly 9,000 m². Appendix Figure A.1 shows the interior appearance of the building. The second and third floors host apparel and fabric wholesalers, while the first floor contains various other types of stores.

The market serves as the central node in northern Vietnam’s garment and textile supply chain. Buyers of apparel wholesalers are primarily small retail-shop owners in Hanoi and surrounding provinces, whereas fabric wholesalers typically supply manufacturers and tailors in Hanoi and surrounding provinces. Similar apparel/fabric wholesale markets are found widely across Asia, including the Zhongda Fabric Market in Guangzhou and the Semba Center Building in Osaka, which feature comparable within-building structures, as well as Dongdaemun in Seoul and Ninh Hiep near Hanoi, which consist of street-level clusters rather than a single enclosed complex.

2.2 Market Governance and Slot Ownership Structure

The market is managed by the Dong Xuan Corporation, which collects monthly fees, maintains common areas, and enforces general rules. Each shop occupies a “slot”—a unit of physical space whose *usage right* is privately transferable. Owners may sell this right to another person for a negotiated payment and must simply report the change in ownerships to the Corporation. Importantly, the Corporation does not mediate transactions nor provide

assistance in matching buyers and sellers. In practice, owners typically advertise availability through signage at the shop or rely on informal personal networks to identify potential buyers.

Owners may also rent out their slots to other sellers, although rental arrangements are relatively uncommon: according to our survey, only about 6% of slots were rented in 2023.

2.3 Historical Context and the Lottery in 1996

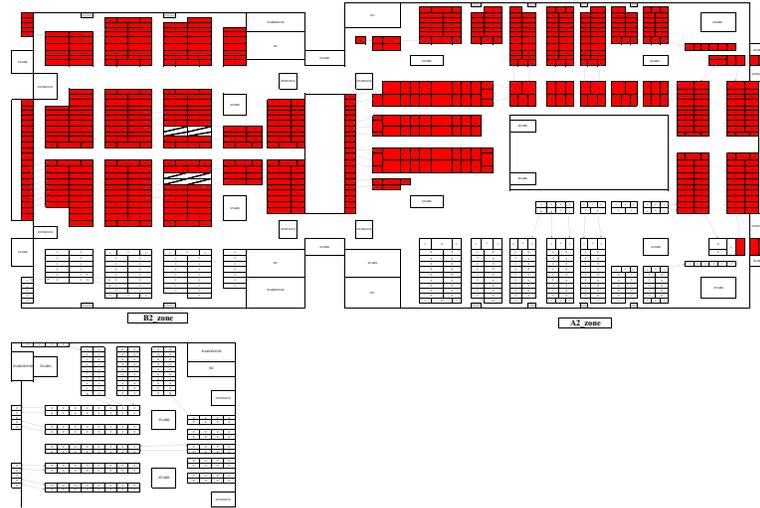
Dong Xuan Market was originally constructed in 1890 during the French colonial period and has remained a landmark of Hanoi’s commercial center. In 1994, a major fire destroyed large portions of the building, forcing merchants to operate temporarily at another site while the market was reconstructed. When the new building neared completion in 1996, the Dong Xuan Corporation faced the challenge of reallocating more than one thousand slots among pre-fire shop owners. To ensure fairness, the Corporation implemented a public lottery in December 1996. Business then resumed in the new building in January 1997. Appendix Figure A.2 presents a local newspaper article reporting the lottery.

According to local newspapers and our interviews with officials of the Corporation, the lottery was implemented through multiple segmented draws. Each shop owner participated in one segment for each slot they had owned prior to the fire. Eligibility for each segment was predetermined based on the characteristics of those pre-fire slots. The segmentation was defined by the intersection of three dimensions of slot characteristics of the reconstructed building:

1. **Floor.** The second floor was the more desirable than the third floor because of higher customer traffic; hence, its lottery was held first for all participants. Participants who could not obtain a slot in that round entered the third-floor lottery.
2. **Fabric/Apparel Zone.** Second floor was divided into *apparel* and *fabric* zones. Figure 1 shows fabric zone in the map. Eligibility for each was determined by the type of products an owner sold before the fire. Switching between apparel and fabric products within each zone was prohibited prior to 2012 when the Corporation abolished this regulation.
3. **Slot Position Rank.** Slots were categorized into three positional ranks based on visibility and accessibility: *best corner* (corner slots facing a main corridor), *second-best corner* (corner slots not facing a main corridor), and *normal* (other slots). Owners who previously occupied best or second-best corner positions before the fire entered

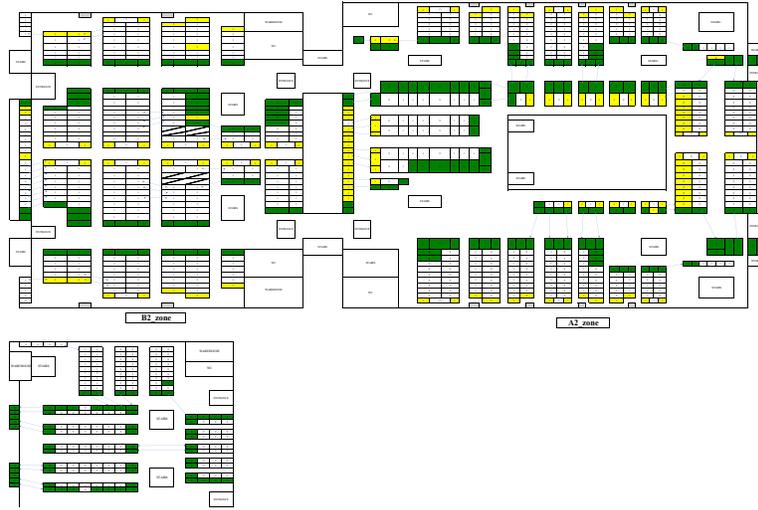
the corresponding lottery for the same rank. Figure 2 shows the locations of the best corner slots.

Figure 1: Apparel and Fabric Zones, 1997–2011



Notes: Slots in the fabric zone are colored in red. The top panel shows the second floor, and the bottom panel shows the third floor.

Figure 2: Slot Position Rank



Notes: Slots identified as the best-corner and second-best-corner slots are colored in green and yellow, respectively. The top panel shows the map of second floor, and the bottom panel shows that of third floor.

2.4 Observations of Clusters in 2023

Two decades after the lottery, visitors to the market observe pronounced spatial clustering of similar products. For example, Figures 3, 4, and 5 show the pictures (taken in 2023) of clusters of shops selling men's pants, women's roomwear, or fabrics for suits. These patterns motivate our central question: how did such clusters emerge?

Figure 3: Cluster of Men's Pants Shops (3F)



Figure 4: Cluster of Women's Roomwear Shops (2F)



Figure 5: Cluster of Fabric-for-Suits Shops (2F)



3 Data

3.1 Survey Design and Coverage

To examine how product clusters emerged after the lottery, we conducted an extensive shop-owner survey between September and October 2023. Enumerators visited each slot inside Dong Xuan Market and interviewed the current owners face-to-face. For owners who were absent, phone interviews were conducted. The survey achieved a response rate of 73.8 percent, which is relatively high for firm surveys.

The survey collected information on slot ownership histories, participation in the 1996 lottery, product types sold in 1994 (before the fire) and subsequently from 1997 onward, as well as business performance and the use of the slot (shop or warehouse). For all slots, the enumerators also photographed product displays to create a visual record of the products in 2023.

In addition, we conducted a short neighborhood survey in which respondents identified the historical products sold by neighboring shops. This procedure enabled us to reconstruct 91 percent of missing product information for 1997 and 2010.

3.2 Product Classification

To standardize product categories, we developed a detailed coding scheme covering 61 distinct product types for apparel and fabrics. Apparel codes are defined by product form and the gender/age-group of end users (e.g., men’s shirts, blouse/shirts for young women, clothing for girls (6-14 years)). Product codes in fabric are classified by intended use (e.g., fabric for men’s suits, blouse, fabric for women’s blouse/shirts, or fabric for women’s roomwear). Figure 6 shows the classification. In the survey, respondents could list multiple product codes that apply. We specifically ask about summer products and winter products, separately. We use summer products in our baseline analysis, although using winter products instead produce similar results.

Figure 6: Product Type Classification (61 Categories)

A1: 0-18 months baby	D1: T-shirt for men
A2: 1.5-5 years girl	D2: Shirts for men
A3: 1.5-5 years boy	D3: Sweaters for men
A4: 6-14 years girl	D4: Other tops for men
A5: 6-14 years boy	D5: Pants for men (other than suits)
B1.1: T-shirt for young ladies	D6: Jeans for men
B1.2: Blouse/Shirts for young ladies	D7: Shorts for men
B1.3: Sweaters for young ladies	D8: Other pants for men
B1.4: Other tops for young ladies	D9: Underwear for men
B1.5: Pants for young ladies (other than suits)	D10: Jacket/coat for men
B1.6: Jeans for young ladies	D11: Formal suits for men
B1.7: Other pants for young ladies	D12: Sports clothes for men
B1.8: Skirts for young ladies (other than suits)	D13: Mitten/scarf for men
B1.9: Dress/One-piece for young ladies	F1.1: Fabrics for men’s shirts
B1.10: Jacket/coat for young ladies	F1.2: Fabrics for men’s pants/suit (other than jeans)
B1.11: Roomwear for young ladies	
B2.1: T-shirt for middle-aged ladies	F2.1: Fabrics for women’s blouse/shirts
B2.2: Blouse/Shirts for middle-aged ladies	F2.2: Fabrics for women’s pants/suit (other than jeans)
B2.3: Sweaters for middle-aged ladies	F2.3: Fabrics for women’s dress
B2.4: Other tops for middle-aged ladies	F2.4: Fabrics for women’s roomwear
B2.5: Pants for middle-aged ladies	F2.5: Fabrics for Ao Dai and traditional formal wear
B2.6: Jeans for middle-aged ladies	
B2.7: Other pants for middle-aged ladies	F3: Fabrics for jeans
B2.8: Skirts for middle-aged ladies	F4: Fabrics for students’ uniform
B2.9: Dress/One-piece/Roomwear for middle-aged ladies	F5: Fabrics for soldiers’ uniform
B2.10: Jacket/coat for young ladies	F6: Fabrics for curtain
B3: Underwear for women	F7: Fabrics for bedding
B4: Ao Dai for women	F8: Lining fabrics
B5: Formal suits for women	
B6: Sunscreen coat for women	F10: Fabrics for worker cloth (excluding office cloth)
B7: Sports clothes for women	F11: Fabrics for others
B8: Mitten/scarf for women	

3.3 Distance Measurement

We use the market’s rectangular grid to measure spatial proximity. Each slot has three immediate neighbors (front, left, right), defined as a distance of 1.² We construct an adjacency matrix defining immediate neighbors for all slots and calculate step distances iteratively; for instance, slots reachable within 3 steps are considered “3-step neighbors.”

3.4 Spatial Maps of Products

Products in 2023

Using these product codes, we produced maps of the main product type for each slot. Figure 7 presents the distribution of the main products in 2023 based on information from the photographs.

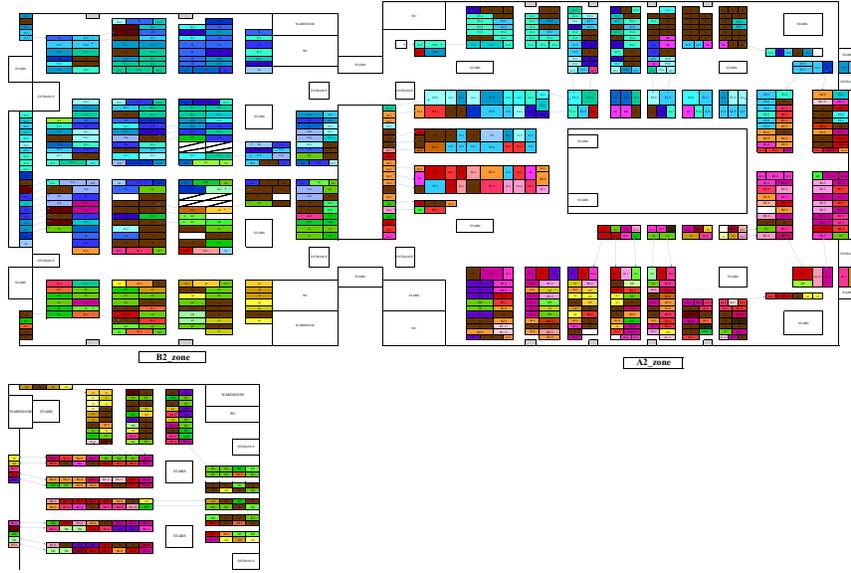
One notable pattern is that stores selling fabric products remain concentrated in the fabric zone (see Figure 1 for fabric zone). However, some slots that were originally designated as fabric zone now sell apparel, particularly along the boundary where fabric and apparel zones meet. These conversions likely occurred after the market corporation relaxed the apparel–fabric zoning regulations in 2012. While this erosion of fabric zone by apparel products near the zone border is interesting in itself, providing additional evidence for the strength of agglomeration externalities, this is not the major focus of our analysis in this paper.

A further salient pattern is the existence of dense clusters of shops selling the same product type (across 61 categories) *within both apparel and fabric zones*. Explaining the formation of these clusters is the central objective of this paper.

Finally, it is important to note that shops typically sell multiple product types, which is a feature not fully reflected in the visual map. For ease of presentation, the map assigns each slot a single main product category, but most shops offer more than one. The mean number of product types per shop has remained nearly constant over the past three decades—1.60 (1994), 1.53 (1997), 1.52 (2010), and 1.53 (2023)—indicating stable product diversification. Accordingly, in our empirical analysis we construct the dataset at the slot (i) \times product (k) level, treating each slot–product pair as an observation.

²Figure A.3 illustrates some examples. Slots in the back are not connected because of walls.

Figure 7: Main Product of Each Slot (2023)



Products initially assigned by the lottery

We next examine the spatial distribution of original products of shops assigned by the 1996 lottery. Because products sold immediately after market reopening in 1997 (measured by summer products in 1997) may already reflect endogenous adjustments, we proxy use 1994 pre-fire products of the shop owner who was assigned to each slot by the lottery. This pre-fire product information is obtained from the shop-owner survey and available for the owners who participated the lottery. For slots that have been transferred from original lottery-assigned owners to new owners, this pre-fire product information is generally missing (unless the original owner continues to operate in 2023 at a different slot and responded to our survey). Among the 1,046 slots in the market, 266 have complete information on the 1994 products of their original owners.

Figure 8 shows the map of 1994 products of the lottery-assigned owner at each slot. The map visually shows that there are no systematic spatial correlation of products within apparel and fabric zones.

Since this 1994 product information is a key exogenous variation we exploit in our empirical analysis, we extensively investigate the nature of sample selection based on availability of this information and conduct robustness checks in following sections.

Figure 8: Hypothetical Product Map Based on 1994 Products of Lottery Winners



4 Empirical Strategy and Sample-Selection and Randomization Checks

4.1 Baseline Specification

Our primary question is whether initial coincidental clustering of similar products leads to the subsequent adoption of those products by neighboring stores. To test this, we estimate the following specification using slot–product level data for each year t :

$$\text{Prod}_{i,k}^t = \beta^t \text{ProdNeigh}(m)_{i,k}^{1994} + \text{StrataProdFE}_{i,k} + \alpha_i^t + \varepsilon_{i,k}^t, \quad (1)$$

where $\text{Prod}_{i,k}^t$ equals 1 if slot i sells product k in year t , and 0 otherwise. The key explanatory variable, $\text{ProdNeigh}(m)_{i,k}^{1994}$, is the share of i 's neighbors within m steps (where m is defined as an integer from 2 to 9) that sold product k before the fire. $\text{StrataProdFE}_{i,k}$ denotes the full set of lottery-segment fixed effects (floor \times apparel/fabric zone \times slot-rank) interacted with product fixed effects. α_i^t is a fixed slot effect, which controls the time-invariant features regarding the number of products at the slot. Standard errors are clustered at the

slot level, and spatial-clustered errors are used as robustness checks.

We estimate Equation (1) separately for four time points: 1994 (pre-lottery), 1997 (immediately after reopening), 2010, and 2023. When this equation is estimated for $t = 1994$, the coefficient β^{1994} provides a test of whether the lottery produced truly random variation in neighbors' initial products. Under random assignment, β^{1994} should be statistically indistinguishable from zero.

For years $t \geq 1997$, the coefficients β^t measure the extent to which initial coincidental concentration of similar products in the local neighborhood increases the subsequent likelihood that a slot specializes in the same product type. A positive and persistent β^t over time indicates that agglomeration externalities are sufficiently strong to induce endogenous clustering.

We mainly focus on the balanced-panel sample that includes slots with pre-fire product information. Summary statistics for the balanced sample are reported in Table 1 (similar statistics for expanded sample are reported in Appendix Table A.2). For years $t \geq 1997$, we additionally estimate the equation with expanded sample which adds slots whose later products are known from neighborhood recall (for 1997 and 2010) and from photographs (for 2023).

Table 1: Summary Statistics: Balanced-Sample Slots with Non-Missing 1994 Product Information

	count	mean	min	max
=1 if slot i sells product k in 1994	16043	0.022	0	1
=1 if slot i sells product k in 1997	15921	0.022	0	1
=1 if slot i sells product k in 2010	15982	0.022	0	1
=1 if slot i sells product k in 2023	16043	0.020	0	1
=1 if slot is best corner	16043	0.35	0	1
=1 if slot is second best corner	16043	0.15	0	1
=1 if slot i is in fabric zone	16043	0.63	0	1
=1 if slot i is on the third floor	16043	0.11	0	1
Share of 2-step neighbors selling product k pre-lottery	13969	0.022	0	1
Share of 3-step neighbors selling product k pre-lottery	15372	0.021	0	1
Share of 6-step neighbors selling product k pre-lottery	15982	0.021	0	1
Share of 9-step neighbors selling product k pre-lottery	16043	0.022	0	1
Number of 2-step neighbors with pre-lottery product info	16043	2.02	0	6
Number of 3-step neighbors with pre-lottery product info	16043	3.36	0	10
Number of 6-step neighbors with pre-lottery product info	16043	8.49	0	22
Number of 9-step neighbors with pre-lottery product info	16043	15.9	1	39

4.2 Sample-Selection Checks

Because the empirical strategy relies on the availability of pre-fire (1994) product information to construct measures of initial coincidental agglomeration, it is important to examine the nature of sample selection and assess whether missing historical information may bias our estimates. We therefore conduct a set of sample-selection checks to examine whether (i) spatial position and basic features of slots predict missingness, and (ii) neighborhood product concentration is correlated with missingness in ways that might bias our estimates. Across all tests, we find no evidence that sample selection threatens our main results.

We begin by comparing the spatial patterns of the full set of slots with those for which 1994 product information is observed. Appendix Figure A.4 plots the distribution of pairwise distances among (i) all slot pairs in the market and (ii) pairs of slots with non-missing 1994 product information. The two distributions closely overlap, indicating that the availability of 1994 information is not systematically related to whether a slot is centrally located or more peripheral within the building.

We next investigate which basic slot characteristics predict missingness. In Appendix Table A.1, we estimate selection regressions where the dependent variable equals one if a slot’s 1994 product information is observed. Average distance to other slots (a proxy for peripheral position) and whether a slot is in the pre-designated fabric zone do not significantly predict missingness. The only attributes associated with significantly lower missingness are location on the second floor and best-corner position. This pattern is consistent with these slots being more accessible from consumers and thus more stable over time. Importantly, in our empirical specification we absorb these characteristics through interactions of lottery-strata fixed effects (floor \times zone \times positional rank) with product fixed effects, ensuring that any differences related to these features do not bias our estimates.

An additional observation to be noted in Appendix Table A.1 is that there are more missingness in slots in the fabric zone that were later converted to apparel after the Corporation relaxed zoning restrictions in 2012. One possibility is that these areas experienced relatively high turnover, leading to greater missingness in pre-fire product information. As a robustness check, we re-estimate our main results excluding these converted slots. The findings remain nearly identical, indicating that this pattern does not affect our results.

We then turn to selection concerns specific to our empirical setting: whether neighborhood-level product concentration predicts missingness or product types. First, if stores surrounded by neighbors selling similar products were more likely to survive and thus be observed, whereas those surrounded by more heterogeneous neighbors were more likely to exit and become unobservable, we could mistakenly infer clustering that arises mechanically from sample selection. We find no such evidence. In Appendix Figure A.5, we show the result of

regressing an indicator for non-missing 1994 information on $\max_k \{\text{ProdNeigh}(m)_{i,k}^{1994}\}$ (measuring the degree of concentration of any product at i) controlling for lottery strata fixed effects. For all range of neighborhood definitions, the correlation is statistically indistinguishable from zero. This implies that stores located in neighborhoods with initially high concentration of a particular product are no more or less likely to have non-missing pre-fire product information than stores surrounded by more mixed neighbors.

Second, a related concern is product-specific survival depending on neighbors' products. Having more neighbors selling product k may help only product k seller to survive and remain observable, but not those selling other products. This selection pattern may lead to spurious spatial correlation among product- k shops. Estimating the baseline equation 1 for $t = 1994$ corresponds to the check against this concern by examining whether the share of neighbors selling product k (as of 1994) predicts the likelihood of observing a product- k seller (as of 1994). As we report in the next main-result section, we find no such relationship, suggesting that neighborhood concentration does not drive product-specific selection into the sample.

Taken together, these sample-selection checks demonstrate that missing 1994 product information does not pose a credible threat to our identification strategy.

4.3 Randomization Check through Simulation

Before turning to the main results, we first assess whether the observed allocation generated by the 1996 lottery is consistent with random assignment of product types. We conduct simulation-based inference following the approach of Duranton and Overman (2005) and Nakajima and Teshima (2020).

For each product k and lottery segment h (defined by the intersection of floor, apparel/fabric zone, and slot-position rank), we identify the set of firms that sold product k in 1994 and were assigned to slots within segment h . Let $n_{k,h}$ denote the number of such firms. We also identify the full set of eligible slots S_h for segment h , including those where 1994 product information is missing.

We make one adjustment: we exclude slots located in the portion of the fabric zone later converted to apparel. These slots have high missingness in pre-fire product information due to ownership turnover (see Section 4.2) and are located on the periphery of the fabric area. Including them in S_h would artificially inflate simulated long-distance pairings relative to the observed data. To maintain consistency throughout this exercise, we also exclude the firms assigned by the lottery to such slots when counting $n_{k,h}$, and we drop these slots when constructing the actual distance distribution for comparison.

We simulate the lottery 1,000 times. In each draw, for every segment h and product k ,

the $n_{k,h}$ firms are randomly reassigned to the slots in S_h . Pooling the simulated allocations for all products and segments, we compute the kernel density distribution of pairwise distances between slots selling the same product (for distances $d = 1, \dots, 40$). From the 1,000 simulated distributions, we construct a 1st–99th percentile confidence interval for each d .

Figure 9: Same-Product Distance over Time vs. Simulated Confidence Interval

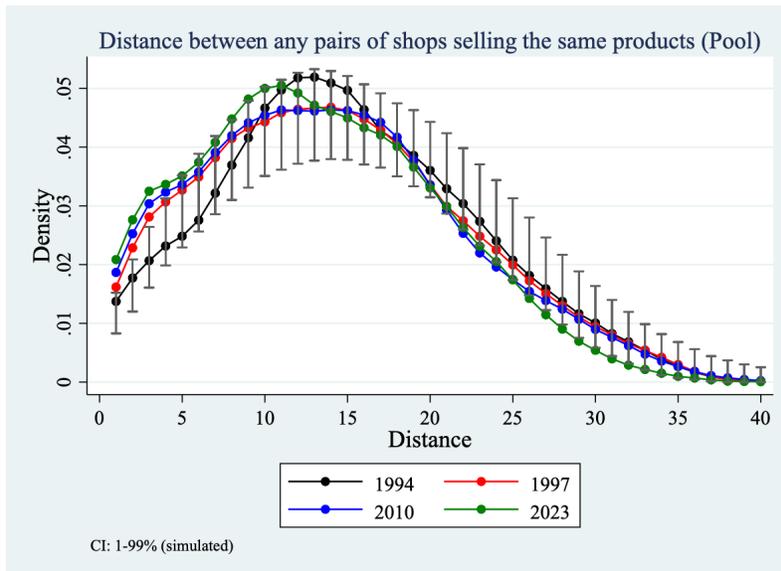


Figure 9 plots the observed 1994 distance distribution (black dots) against this simulated confidence interval. The observed curve lies within the confidence interval, indicating that the spatial distribution of products implied by the lottery outcome is statistically indistinguishable from random assignment.

We then repeat this exercise for 1997, 2010, and 2023 using the balanced-panel sample of slots for which 1994 product information is observed. As shown in Figure 9, the density of short-distance same-product pairs rises sharply above the upper confidence band beginning in 1997 and continues to increase through 2010 and 2023. These patterns indicate that product clustering emerged rapidly after reopening and strengthened over time, reflecting endogenous adjustments rather than mechanical persistence of pre-fire locations.

As an additional robustness check, we repeat the simulations including the slots later converted from fabric to apparel. The result reported in Appendix Figure A.6 remains qualitatively unchanged for distances below approximately ten steps, although confidence intervals widen at longer distances—well beyond the spatial scale at which clustering occurs in this setting. This confirms that the exclusion of these slots does not drive our conclusions.

Overall, the simulation analysis validates the identifying assumption: any systematic spatial clustering observed after the lottery arises from post-assignment adjustments rather

than pre-existing sorting or location preferences.

5 Main Results

5.1 Main Effects of Initial Agglomeration

Table 2 reports the baseline results of estimating equation 1 where neighborhoods are defined by three-step proximity. The dependent variable is an indicator for whether slot i sells product k in year t . “Share of neighbors selling k in 1994” is the share of i ’s neighbors within three steps that sold product k before the fire (i.e., $\text{ProdNeigh}(3)_{i,k}^{1994}$). Panel A restricts the sample to the balanced panel of slots with non-missing 1994 products, while Panel B includes all slots with available outcomes in each year.

Table 2: Effect of Initial Coincidental Agglomeration (3-Step Neighborhoods)

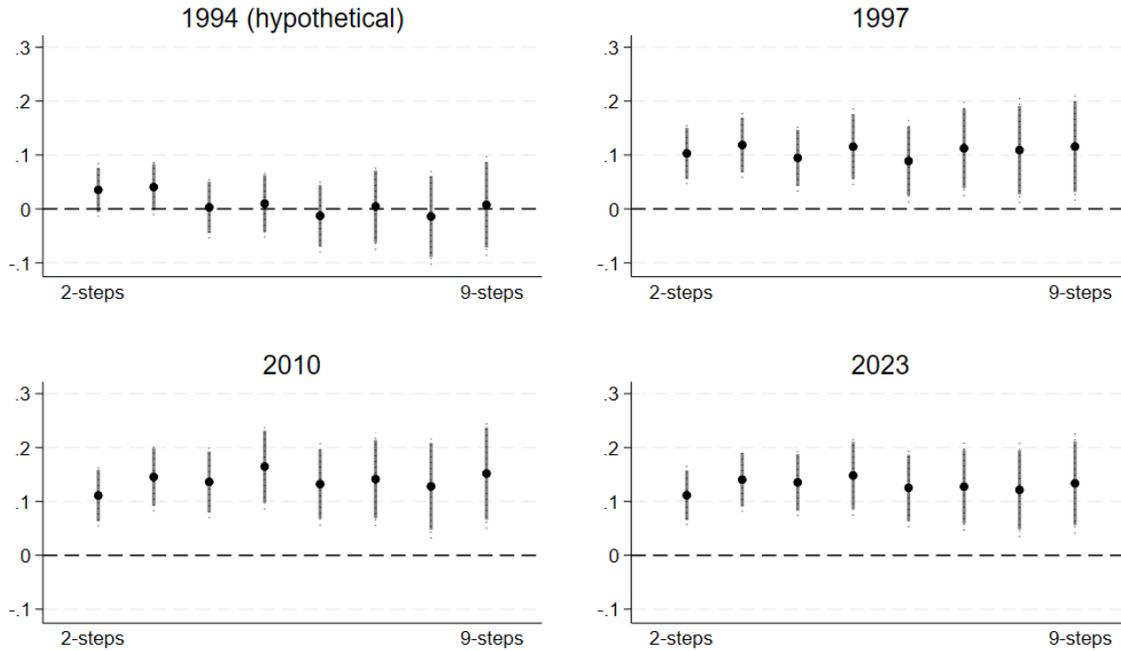
	(1)	(2)	(3)	(4)
	1994	1997	2010	2023
	product	product	product	product
<i>Panel A: Using sample of slots with non-missing 1994 info</i>				
Share of neighbors selling k in 1994	0.039	0.118***	0.146***	0.141***
	(0.026)	(0.031)	(0.033)	(0.030)
Observations	15372	15311	15311	15372
<i>Panel B: Using sample of all available slots with info in each year</i>				
Share of neighbors selling k in 1994		0.100***	0.117***	0.099***
		(0.016)	(0.015)	(0.013)
Observations		50386	55327	56486

Column (1) confirms that before the fire, there was no spatial correlation between a slot’s own product and that of its neighbors, consistent with random assignment. This result also alleviates concerns regarding product-specific sample selection driven by neighbors’ products (as discussed in Section 4.2). Columns (2)–(4) show that immediately after reopening in 1997, stores located near same-product neighbors became significantly more likely to sell that product. The magnitude of this effect rises by 2010 and remains stable through 2023, indicating persistent and strengthening clustering over time.

Figure 10 visualizes the coefficients across years and alternative neighborhood definitions ($m = 2, \dots, 9$). All specifications yield near-zero estimates for 1994, positive coefficients

in 1997, and larger, statistically significant coefficients thereafter. The pattern is robust to replacing the neighborhood definition to spatially-clustered standard errors (Figure ??).

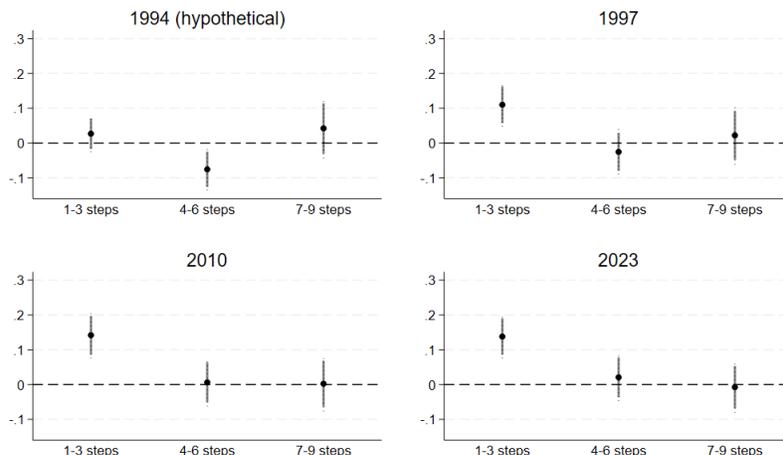
Figure 10: Coefficients by Year and Neighborhood Size



5.2 Spatial Range of Externalities

To identify how far the externalities extend, we decompose neighborhoods into concentric rings: 1–3 steps, 4–6 steps, and 7–9 steps. Including these variables jointly shows that only the closest neighbors matter as the effects decline sharply with distance (Figure 11). This pattern suggests that the relevant interactions occur within a very local spatial scale, roughly equivalent to a few adjacent slots.

Figure 11: Effects by Distance Rings (1–3, 4–6, 7–9 Steps)



In summary, the positive and persistent relationship between initial coincidental clustering and subsequent product choices implies that local agglomeration externalities are strong enough to overcome the competitive disadvantages of proximity and adjustment costs.

6 Mechanisms

This section explores the underlying processes through which the initially random allocation of store slots evolved into persistent product clusters. We focus on two complementary dimensions. First, we study the *channels of adjustment* through which spatial reorganization occurred: that is, whether clustering emerged from existing owners switching their product lines or from ownership transfers that changed the identity of sellers. Second, we examine the *economic mechanisms of agglomeration externalities*: that is, whether they operate primarily through demand-side forces such as customer attraction and shared foot traffic, or through supply-side knowledge diffusion across sellers.

6.1 Channels of Product Adjustment

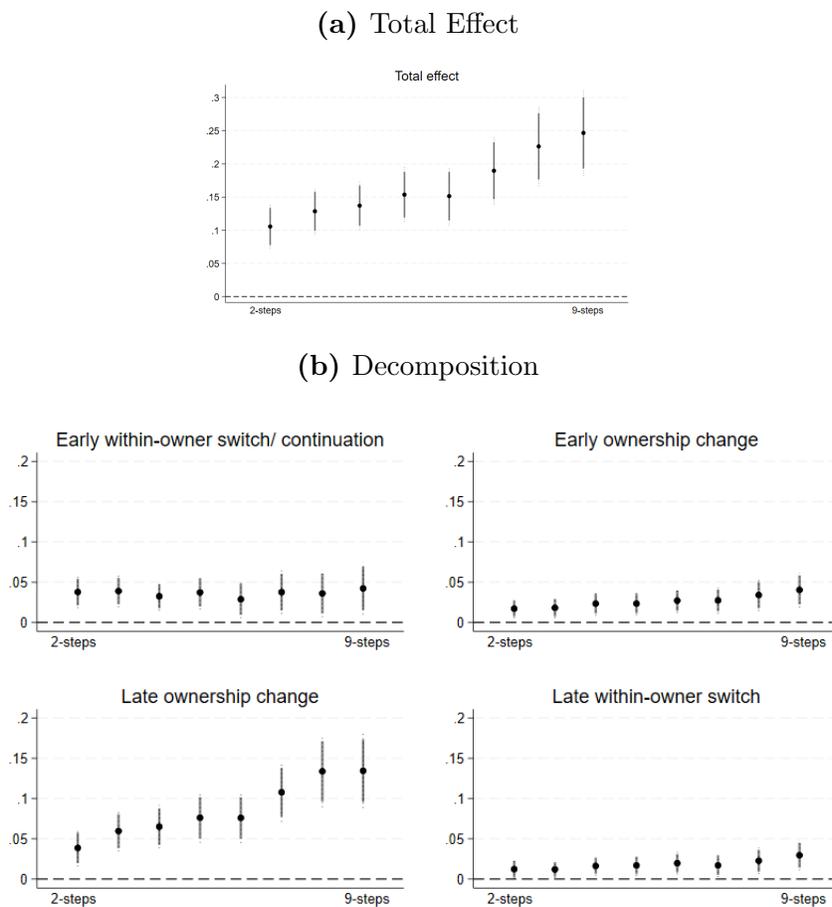
Clustering can arise either through the behavioral responses of incumbent owners or through market transactions that reallocate slots to new owners. To identify these channels, we begin by collapsing the time dimension and defining a binary variable $\text{Total}_{i,k} = 1$ if product k was sold in slot i at any time between 1997 and 2023, and 0 otherwise. Re-estimating Equation (1) with this cumulative outcome captures the total response to initial coincidental clustering.

Here, we include slots for which the 1994 product information of the initial owner is

missing (i.e., not focusing on the balanced panel sample). In other words, we retain a slot as long as we observe its ownership and product history in some years and observe the treatment variable (initial neighborhood product share). This is important for capturing the effect of ownership transfers: even when 1994 product information for the original owner is missing, we often observe subsequent owners in later years and can track the products they sold up to 2023.

As reported in Figure 12 Panel (a), the estimated coefficient rises steadily with neighborhood size, indicating that the cumulative influence of neighboring products compounds over spatial distance.

Figure 12: Decomposition of Product Changes



We then decompose $Total_{i,k}$ into four mutually exclusive cases that distinguish the timing of adjustments and whether they occurred within or across owners:

1. **Early within-owner switch/continuation:** product k sold in 1997 by the same owner who obtained the slot in the lottery;

2. **Early ownership transfer:** product k sold in 1997 by a new owner;
3. **Late ownership transfer:** product k first appearing between 1998 and 2023 following an ownership change;
4. **Late within-owner switch:** product k newly introduced after 1998 by the original owner.

Appendix Table A.3 reports the frequencies of these categories.

The results presented in Figure 12, Panel (b), show that both types of adjustment contributed to the eventual clustering of products, but their timing and intensity differed. Ownership transfers after 1998 account for the largest share of the total effect, suggesting that part of the spatial reorganization gradually through entries of owners who sold products similar to their neighbors.

At the same time, a nontrivial share of the response, particularly in the first year after reopening (*within-owner switching*) are made by incumbents quickly adapted their product lines to their stores. To better understand the nature of within-owner adjustments, we track the evolution of product portfolios over time within owners. For each slot, we identify whether a given product type was (i) continued, (ii) newly added, or (iii) dropped relative to the owner’s 1994 product set. We then estimate variants of Equation (1) using these outcomes as dependent variables. Appendix Figure A.8 plots the resulting coefficients. Slots surrounded by a higher initial concentration of product k are significantly more likely to add that product to their line-up and less likely to drop it 1997.

Taken together, the evidence indicates that the gains from local agglomeration externalities were substantial enough to offset both the costs of adjusting product lines and the frictions associated with ownership transfers, even in the presence of intensified local competition.

6.2 Demand- versus Supply-Side Externalities

Finally, we examine the economic mechanisms through which agglomeration externalities operate. The literature on agglomeration externalities in service sector have emphasized two broad channels: demand- and supply-side channels. On the *demand side*, clustering may arise from “shopping externalities,” whereby co-location increases the overall attractiveness of an area to customers. Early models by Wolinsky (1983), Dudey (1990), and Konishi (2005) show that when consumers face search or transportation costs, firms have incentives to locate near competitors to exploit customer pooling and comparison shopping. Later theoretical extensions, such as Fischer and Harrington Jr. (1996) and Takahashi (2013), demonstrate

that the coexistence of product variety and imperfect information can sustain equilibrium clustering even under price competition. Empirically, Shoag and Veuger (2018) and Nakajima and Teshima (2020) provide evidence that demand-side agglomeration effects are substantial: the closure of a large retail store negatively affects its neighbors, and lottery-based location allocation in Tokyo’s fish market generates positive spillovers among proximate sellers.

On the *supply side*, spatial proximity may facilitate information exchange and learning among similar firms. Previous works focusing on knowledge-intensive sectors have found evidence for such spillovers across firms within narrow geographic areas (Arzaghi and Henderson, 2008; Atkin et al., 2022). In this view, even in our setting, clusters may form because neighboring stores observe and imitate successful practices such as product designs, suppliers, or input combinations, creating local knowledge spillovers that enhance productivity. These two channels are not mutually exclusive, and distinguishing them empirically is challenging because both generate similar spatial correlations in firm behavior.

We exploit the presence of multi-slot owners in our setting to infer which channel is more salient. If supply-side learning were the dominant mechanism, an owner operating multiple slots in different areas should transfer information gained in one location to another. Formally, we estimate: for $t = 1997$ and 2023,

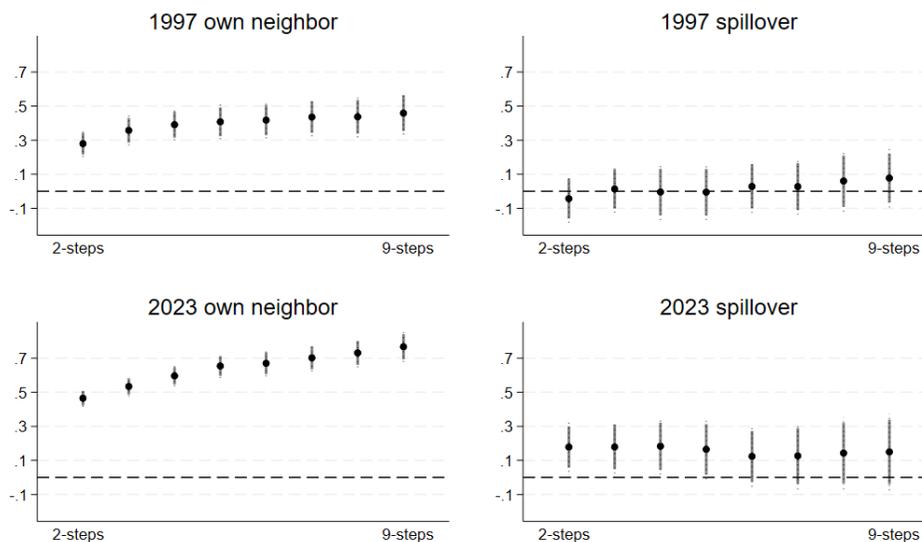
$$\text{Prod}_{i,j,k}^t = \beta_1^t \text{ProdNeigh}(m)_{i,j,k}^t + \beta_2^t \text{ProdNeigh}(m)_{-i,j,k}^t + \text{StrataProdFE}_{i,k} + \alpha_i^t + \varepsilon_{i,j,k}^t,$$

where $\text{ProdNeigh}(m)_{i,j,k}^t$ denotes the share of slot i ’s local neighbors selling product k in year t , and $\text{ProdNeigh}(m)_{-i,j,k}^t$ represents the share of neighbors surrounding the owner’s other slot(s) in year t . For owners who had only a single slot in year t , the latter variable takes 0. If knowledge spillovers were important, both β_1 and β_2 would be positive and similar in magnitude.

Ideally, we would implement this test using pre-lottery information from 1994 to measure neighborhood product information (i.e., using $\text{ProdNeigh}(m)_{i,j,k}^{1994}$ and $\text{ProdNeigh}(m)_{-i,j,k}^{1994}$). However, very few lottery participants who owned multiple slots from lottery are observed in our survey data with identifiable pre-lottery product information. This makes a 1994-based analysis statistically infeasible. We therefore rely on ownership and product information from 1997 and 2023 (combining neighborhood survey and product photographs with owner survey), where we observe more multi-slot owners with product information (around 7–8 percent of all owners, as reported in Appendix Table A.5). While these later data inevitably reflect some endogenous adjustment through product changes and ownership transfers, we still believe that they provide suggestive evidence to assess the relative strength of within- and cross-location spillovers.

Empirically, we find that β_1 is large and statistically significant, while β_2 is close to zero in 1997 (Figure 13). Even for 2023, when the endogenous location choices by multi-slot owners are likely to positively bias the results, the same patterns are observed. This asymmetry between own- and cross-slot spillovers indicates that agglomeration benefits are highly localized and linked to the physical neighborhood of each slot rather than transferable knowledge at the firm level. This evidence favors a *demand-driven* interpretation: clustering emerges because sellers benefit from shared customer flows and collective visibility, rather than from knowledge diffusion across firms.

Figure 13: Neighbor vs. Cross-Slot (Same-Owner) Spillovers



7 Conclusion

This paper provides causal evidence that retail and service clusters can emerge endogenously from localized agglomeration externalities. Using the natural experiment of a lottery-based allocation of wholesale market slots in Hanoi, we show that an initially random distribution of product types evolved rapidly into distinct, persistent clusters. Areas with greater coincidental pre-lottery concentration of a product were substantially more likely to see neighboring shops adopt to the same product within a year, with effects enduring for more than two decades.

Our findings highlight several mechanisms underlying this process. First, clustering occurred through both rapid within-owner product switching and slower ownership transfers, indicating that firms' private benefits from agglomeration externalities were large enough to

overcome these adjustment costs, in addition to possible price competition. Second, spillovers across distant slots owned by the same firm were limited, implying that agglomeration externalities are primarily demand-driven rather than supply-side knowledge diffusion.

The results have implications for urban and industrial policy. They suggest that self-organization of retail clusters can occur even without centralized coordination, and that overly restrictive zoning or displacement policies may hinder rather than promote efficient spatial organization. Conversely, urban interventions that foster accessibility and customer flows may magnify natural agglomeration forces. More broadly, the evidence contributes to understanding how micro-level interactions among service establishments shape the spatial structure of firms' locations and product choices.

References

- Arzaghi, Mohammad and J. Vernon Henderson**, “Networking off Madison Avenue,” *Review of Economic Studies*, 2008, 75 (4), 1011–1038.
- Atkin, David, M. Keith Chen, and Anton Popov**, “The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley,” *NBER Working Paper*, 2022.
- Baum-Snow, Nathaniel, Nicolas Gendron-Carrier, and Ronni Pavan**, “Local productivity spillovers,” *American Economic Review*, 2024, 114 (4), 1030–1069.
- Benmelech, Efraim, Nittai Bergman, Anna Milanez, and Vladimir Mukharlyamov**, “The agglomeration of bankruptcy,” *The Review of Financial Studies*, 2019, 32 (7), 2541–2586.
- Bernstein, Shai, Emanuele Colonnelli, Xavier Giroud, and Benjamin Iverson**, “Bankruptcy spillovers,” *Journal of Financial Economics*, 2019, 133 (3), 608–633.
- Combes, Pierre-Philippe, Duranton, Gobillon, Diego Puga, and Sébastien Roux**, “The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection,” *Econometrica*, 2012, 80 (6), 2543–2594.
- Dudey, Marc**, “Competition by Choice: The Effect of Consumer Search on Firm Location Decisions,” *American Economic Review*, 1990, 80 (5), 1092–1104.
- Duranton, Gilles and Henry G. Overman**, “Testing for Localisation Using Micro-Geographic Data,” *Review of Economic Studies*, 2005, 72 (4), 1077–1106.

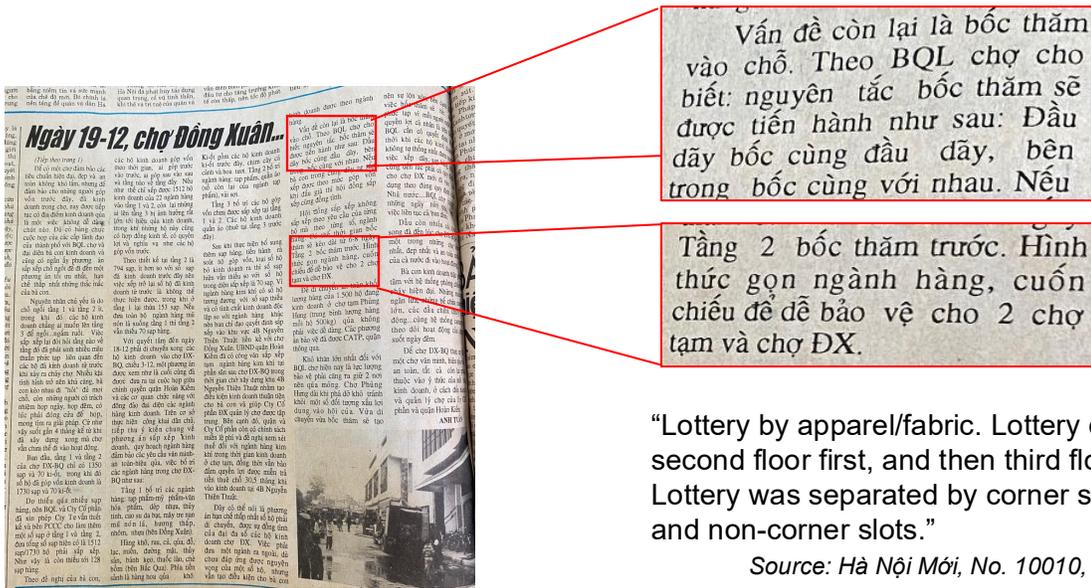
- Ellison, Glenn, Edward L. Glaeser, and William R. Kerr**, “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns,” *American Economic Review*, June 2010, *100* (3), 1195–1213.
- Fischer, Jeff E. and Joseph E. Harrington Jr.**, “Product Variety and Firm Agglomeration,” *RAND Journal of Economics*, 1996, *27* (2), 281–309.
- Konishi, Hideo**, “Concentration of Competing Retail Stores,” *Journal of Urban Economics*, 2005, *58* (3), 488–512.
- Koster, Hans RA, Ilias Pasidis, and Jos van Ommeren**, “Shopping externalities and retail concentration: Evidence from Dutch shopping streets,” *Journal of Urban Economics*, 2019, *114*, 1–29.
- Leonardi, Marco and Enrico Moretti**, “The Agglomeration of Urban Amenities: Evidence from Milan Restaurants,” *American Economic Review: Insights*, 2023, *5* (2), 141–57.
- Marshall, Alfred**, *Principles of Economics*, 8th edition ed., London: Macmillan, 1920.
- Nakajima, Kentaro and Kensuke Teshima**, “Identifying Neighborhood Effects among Firms: Evidence from Location Lotteries of the Tokyo Tsukiji Fish Market,” 2020. RIETI discussion paper no. 18-E-044.
- Shoag, Daniel and Stan Veuger**, “Shops and the City: Evidence on Local Externalities and Local Government Policy from Big-Box Bankruptcies,” *Review of Economics and Statistics*, 2018, *100* (3), 440–453.
- Takahashi, Takaaki**, “Agglomeration in a city with choosy consumers under imperfect information,” *Journal of Urban Economics*, 2013, *76*, 28–42.
- Vitali, Anna**, “Consumer Search and Firm Location: Theory and Evidence from the Garment Sector in Uganda,” Working Paper, IGC 2024.
- Wolinsky, Asher**, “Retail Trade Concentration Due to Consumers’ Imperfect Information,” *Bell Journal of Economics*, 1983, *14* (1), 275–282.

Appendix

Figure A.1: Interior of Dong Xuan Market



Figure A.2: Local Newspaper Report on the 1996 Lottery



“Lottery by apparel/fabric. Lottery on second floor first, and then third floor. Lottery was separated by corner slots and non-corner slots.”

Source: Hà Nội Mới, No. 10010, 1996.

Figure A.3: Measuring Distance between Slots

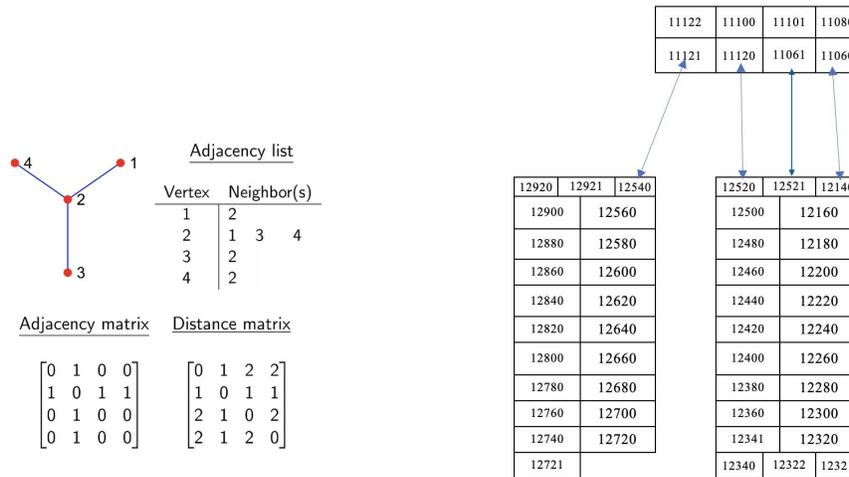


Figure A.4: Distribution of Pairwise Distances: All vs. Non-Missing 1994 Information

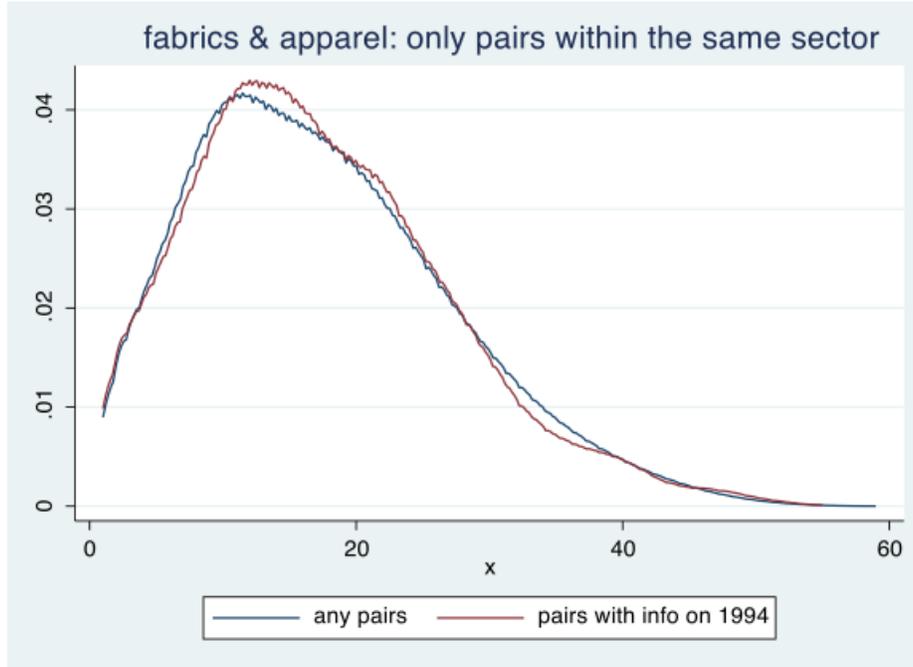


Table A.1: Sample-Selection Checks: Basic feature of slots with non-missing information for 1994 pre-fire products

	Non-missing 1994 information		
Best corner slot	0.156***	0.156***	0.158***
	(0.033)	(0.033)	(0.033)
Second-best corner slot	0.040	0.039	0.048
	(0.042)	(0.042)	(0.042)
Fabric area	-0.017	-0.012	0.019
	(0.033)	(0.035)	(0.037)
3F	-0.133***	-0.119**	-0.108**
	(0.037)	(0.051)	(0.050)
Average distance to other slots		0.001	0.001
		(0.002)	(0.002)
Fabric area slot eroded by apparel			-0.104***
			(0.037)
Observations	1046	1046	1046
Mean	0.229	0.229	0.229

Figure A.5: Sample-Selection Checks: Correlation of neighborhood initial product concentration and missingness

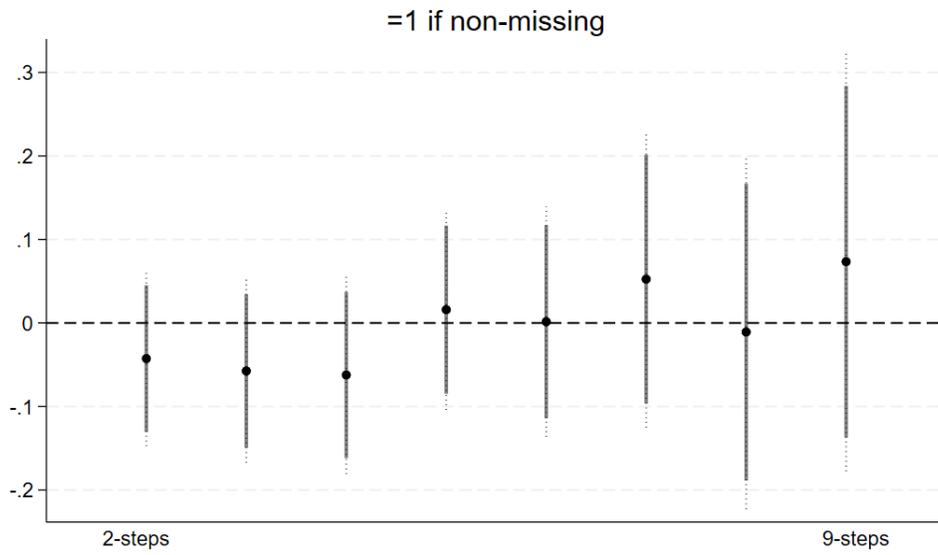


Figure A.6: Distance Confidence Interval (No-Area-Restriction)

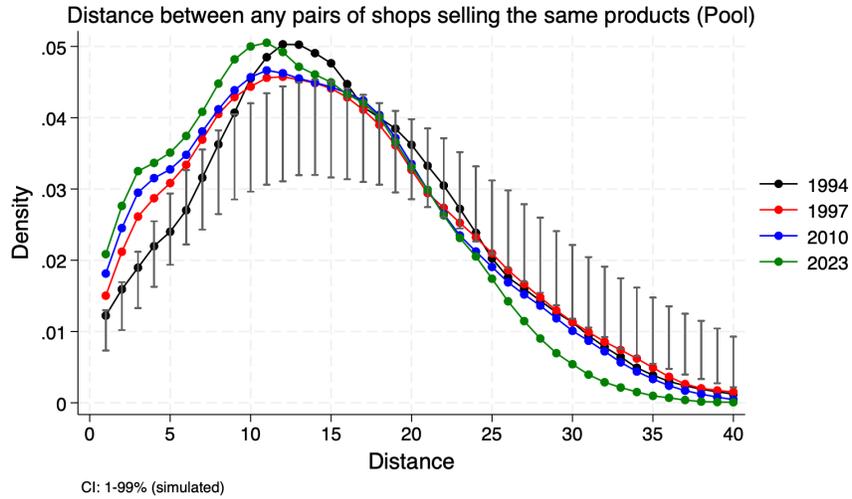


Table A.2: Summary Statistics (Expanded Sample)

	count	mean	min	max
=1 if slot i sells product k in 1994	16043	0.022	0	1
=1 if slot i sells product k in 1997	53924	0.024	0	1
=1 if slot i sells product k in 2010	61305	0.021	0	1
=1 if slot i sells product k in 2023	63074	0.019	0	1
=1 if slot is best corner	63806	0.24	0	1
=1 if slot is second best corner	63806	0.13	0	1
=1 if slot i is in fabric zone	63806	0.59	0	1
=1 if slot i is on the third floor	63806	0.19	0	1
Share of 2-step neighbors selling product k pre-lottery	49288	0.022	0	1
Share of 3-step neighbors selling product k pre-lottery	57157	0.022	0	1
Share of 6-step neighbors selling product k pre-lottery	63379	0.022	0	1
Share of 9-step neighbors selling product k pre-lottery	63806	0.022	0	1
Number of 2-step neighbors with pre-lottery product info	63806	1.68	0	7
Number of 3-step neighbors with pre-lottery product info	63806	2.85	0	11
Number of 6-step neighbors with pre-lottery product info	63806	7.79	0	23
Number of 9-step neighbors with pre-lottery product info	63806	15.0	1	41

Figure A.7: Spatial-Clustering Robustness

	(1)	(2)	(3)	(4)
	1994 product	1997 product	2010 product	2023 product
<i>Panel A: using sample of slots with non-missing info 1994</i>				
Share neighbors' selling the same products (1994)	0.039 (0.025)	0.13*** (0.029)	0.15*** (0.030)	0.15*** (0.028)
N	14986	14868	14868	14986
<i>Panel B: using sample of all available slots with info in each year</i>				
Share neighbors' selling the same products (1994)		0.11*** (0.015)	0.13*** (0.014)	0.11*** (0.013)
N		49855	53572	52510

Table A.3: Summary Statistics for Ownership/Switching Decomposition

	count	mean
Total	50264	0.025
Early within-owner product switch	50264	0.0071
Early ownership change	50264	0.0024
Late ownership change	50264	0.013
Late within-owner product switch	50264	0.0028

Table A.4: Summary Statistics for Product Switching

	count	mean
=1 if slot i added product k that was unsold pre-lottery by year 1997	16043	0.0060
=1 if slot i added product k that was unsold pre-lottery by year 2010	16043	0.0072
=1 if slot i added product k that was unsold pre-lottery by year 2023	16043	0.0079
=1 if slot i continued product k that was sold pre-lottery in year 1997	16043	0.016
=1 if slot i continued product k that was sold pre-lottery in year 2010	16043	0.014
=1 if slot i continued product k that was sold pre-lottery in year 2023	16043	0.011
=1 if slot i dropped product k that was sold pre-lottery by year 1997	16043	0.0053
=1 if slot i dropped product k that was sold pre-lottery by year 2010	16043	0.0069
=1 if slot i dropped product k that was sold pre-lottery by year 2023	16043	0.0098

Figure A.8: Within-Owner Product Additions and Drops

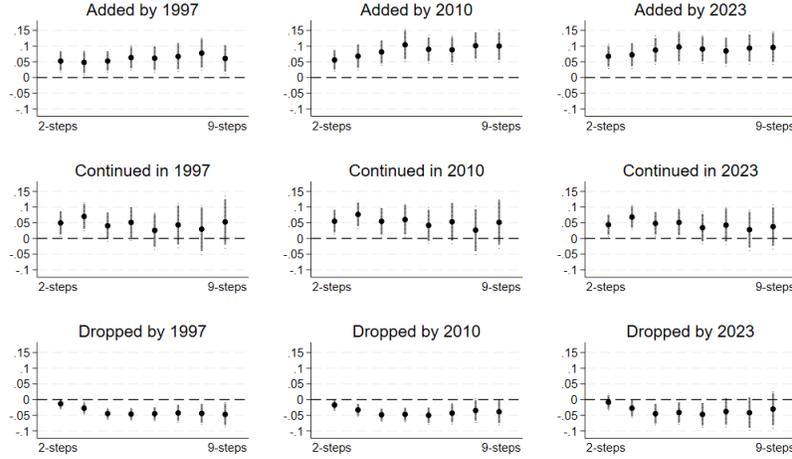


Table A.5: Summary Statistics for Cross-Slot Spillover Analysis

	count	mean	min	max
Share of 2-step neig of -i slots (based on 1997 ownership) selling k in 1997	2013	0.022	0	1
Share of 3-step neig of -i slots (based on 1997 ownership) selling k in 1997	2013	0.023	0	1
Share of 6-step neig of -i slots (based on 1997 ownership) selling k in 1997	2135	0.023	0	1
Share of 9-step neig of -i slots (based on 1997 ownership) selling k in 1997	2135	0.023	0	0.5
=1 if slot i has no -i slots (based on 1997 ownership) in 2-steps	27450	0.93	0	1
=1 if slot i has no -i slots (based on 1997 ownership) in 3-steps	27450	0.93	0	1
=1 if slot i has no -i slots (based on 1997 ownership) in 6-steps	27450	0.92	0	1
=1 if slot i has no -i slots (based on 1997 ownership) in 9-steps	27450	0.92	0	1
Share of 2-step neig of -i slots (based on 2023 ownership) selling k in 2023	4392	0.017	0	0.8
Share of 3-step neig of -i slots (based on 2023 ownership) selling k in 2023	4392	0.018	0	0.6
Share of 6-step neig of -i slots (based on 2023 ownership) selling k in 2023	4392	0.018	0	0.6
Share of 9-step neig of -i slots (based on 2023 ownership) selling k in 2023	4392	0.018	0	0.6
=1 if slot i has no -i slots (based on ownership in 2023) in 2-steps	55693	0.92	0	1
=1 if slot i has no -i slots (based on ownership in 2023) in 3-steps	55693	0.92	0	1
=1 if slot i has no -i slots (based on ownership in 2023) in 6-steps	55693	0.92	0	1
=1 if slot i has no -i slots (based on ownership in 2023) in 9-steps	55693	0.92	0	1