Defensive Specialization: Theory and Evidence from Mexico's Retail Sector

Tommaso Bondi
Cornell University

Luís Cabral
New York University and *CEPR*

Shreya Kankanhalli
Cornell University

Miguel Angel Talamas Marcos
Inter-American Development Bank

January 15, 2026

Abstract. Large companies – both offline and online – are increasingly dominating the retail industry, casting doubt on the future of independent, local retailers. Is this the end of local retail as we know it? What product strategy should local stores follow to attenuate this negative impact? We show that, in response to increasing penetration of large chains, local retailers – and especially smaller ones – optimally follow a strategy of defensive specialization. Intuitively, the arrival of large chains hurts all independent local stores, but it disproportionately harms general stores that sell multiple categories. Our empirical evidence, built on microdata from retail censuses and surveys from Mexico, confirms the predictions of the theoretical model: specialty stores are better able to cope with the shock of large-chain entry. Moreover, when the size disadvantage is large enough, the entry of a large chain induces a shift toward greater specialization in the traditional retail sector, both through an increase in the absolute number and market share of specialty stores and through changes in the product offerings within local stores towards more specialization. These findings challenge conventional wisdom on small store survival in both developing and developed economies and add critical nuance to Anderson's celebrated "Long Tail" theory: small retailers stock niche product categories not because of infinite capacity, but precisely because of their capacity constraints.

1. Introduction

The decline of neighborhood stores amid the rise of retail giants has attracted substantial interest from academics and policymakers alike. Across the globe, the rapid proliferation of chains, big-box retailers, and online platforms has fueled a narrative often described as the "retail apocalypse."[1] Yet this forecast obscures a critical heterogeneity: while many local retailers exit, others adapt and endure. This observation raises fundamental strategic questions: Is this truly the end of local retail as we know it? Which types of local stores are most vulnerable to the rise of retail giants? And, most importantly, what product strategy allows capacity-constrained incumbents to attenuate this negative impact?

We address these questions theoretically and empirically by proposing and testing the optimality of *defensive specialization*: responding to a larger competitor's entry by reducing the number of product categories (narrowing scope) while increasing variety within the remaining categories (increasing depth). We show that defensive specialization is optimal only for incumbents with a large enough capacity disadvantage.

We formalize this intuition through a model of competition between a chain store and a local incumbent. The chain enjoys a significant capacity advantage, consistent with scale economies documented in Jia (2008) and Holmes (2011). The local store chooses between remaining a *general store* (a "miniature" version of the entrant) and pivoting to a *specialty store* (effectively an aisle of the entrant).

Crucial to our contribution is distinguishing *specialization* from *differentiation*. Differentiation implies offering unique products not carried by the entrant. Specialization entails narrowing focus and expanding depth, *even if the entrant sells every product in that category*. To make this distinction as stark as possible, our baseline model assumes the large store offers *all* products, rendering differentiation impossible, isolating specialization as the strategic lever.

Our central result is that optimal assortment response depends critically on the incumbent's *capacity disadvantage*. When the size gap is small, remaining a generalist is optimal. As the gap widens, the generalist strategy collapses: by pivoting to a specialist, the incumbent sacrifices breadth but captures a higher share within a specific niche, increasing total profits.

While our baseline model is deliberately parsimonious, we extend it to study complementarities with pricing, investment in (and, therefore, differentiation through) offline amenities, asymmetric product categories, eclectic consumers, and richer competition including local

1. See, for example, Fox5NY: "Retail apocalypse: How e-commerce slowly killed the neighborhood retail store over the past decade".

stores competing with each other.

We empirically validate these predictions in Mexico's food retail industry by estimating causally how retailers in the traditional sector respond to the massive expansion of convenience and supermarket chains. This setting provides an ideal laboratory: the traditional sector is characterized by the prevalence of small generalist stores (i.e., neighborhood shops) and specialty stores (e.g., bakeries, tortillerias, butchers, and produce stores). Our empirical analysis focuses on specialization. While establishments in the traditional sector (generalists and specialists) may also differentiate from chains by offering, for example, informal credit, community relationships, and physical proximity, we test the optimality of defensive specialization beyond these potential differentiation strategies.

The analysis presents measurement and identification challenges, which we overcome through rich data from multiple sources. First, we leverage two decades of Economic Census microdata from Mexico's National Institute of Statistics and Geography (INEGI), encompassing all physical establishments regardless of size or formality status. Second, we utilize consistent enumerator classifications of store types over time. Third, to address the endogeneity of chain entry location, we employ an instrumental variable strategy introduced in Talamas Marcos (2024). The instrument interacts cross-sectional variation in neighborhood suitability for chains (prevalence of wide streets) with time-series variation in the expected operating costs of chains based on the chain's regional expansion patterns.

The time-series variation follows the logic of Jia (2008) and Holmes (2011) where chains enjoy regional economies of scale by opening stores in nearby cities that share costs (e.g., logistics, overhead, marketing). Since these cost advantages are specific to the chain that opens the stores in nearby cities, the exclusion restriction is plausibly satisfied: traditional-sector stores in markets suitable for chains are affected by a chain's openings in nearby cities only through an increase in the probability that the chain will enter their market. In the Empirical Section, we discuss in detail potential threats to identification and to the validity of the exclusion restriction. We adapt this instrument to separately estimate effects of entry by convenience chains and supermarkets, directly testing our predictions concerning capacity disadvantage magnitude.

Fourth, because census data captures market-level specialization through entry, survival, and exit, but not within-firm specialization in the product offering, we supplement the analysis with two waves of primary surveys of over 500 retail stores in Mexico City. This primary data allows us to look inside the "black box" of the firm, documenting explicit assortment adjustments and the managerial reasoning behind them.

Our empirical results provide robust and causal support for our theoretical model. While the entry of a chain hurts both general and specialty stores, the adverse effects on profits and revenue are larger for general stores. These effects are also increasing in the proximity and the size of the entrant. At the market level, we find that the entry of a chain leads to a significant increase in the share and number of specialty stores. Consistent with our capacity disadvantage mechanism, the magnitude of this effect scales with the entrant's size: the entry of a supermarket drives a 29 percentage point increase (104%) in the share of specialized retailers, compared to a 4.7 percentage point increase (17%) from a convenience chain. Furthermore, we document that within the traditional sector, the entry of a chain leads to a rise in the share of exits by general stores and in the share of entries by specialty stores.

At the store level, our survey data reveals a notable divergence in behavior between smaller and larger local stores that further validates our theory. We find that chain entry significantly increases specialization among *smaller* local stores–driving a 0.28 standard deviation increase on an LLM-based specialization index constructed from managers' reported assortment changes–while larger local stores show negligible effects. This is remarkably aligned with our prediction that defensive specialization is the optimal response specifically for those facing a severe capacity disadvantage.

Although our primary empirical analysis focuses on Mexican retail, U.S. data reveals a similar pattern. Even amid rapid retail consolidation – with the top 20 food retailers increasing market share from 35% to 65.1% between 1990 and 2019 – small specialty retail stores have expanded. Between 2019 and 2024, small specialty retail businesses grew at 6.5% annually, reaching 264,251 establishments, while independent grocery stores in states like Iowa and Nebraska saw sales decline 15–30%.[2]

Mark Cohen, former Director of Retail Studies at Columbia Business School, describes a similar structural shift in a very different retail market, the one for books:

> *There is a tremendous resurgence of local stores, but these have relevance because they're not trying to be all things to all people. They're picking a genre or curating an assortment that appeals to a local customer. [Amazon's] unmatchable scale is liberating for booksellers: it lets them focus on a particular aesthetic rather than stocking items people need but don't get excited about.*

Similarly, Alter (2020) documents increasing diversity in US bookstores, consistent with niche specialization.[3] Taken together, these patterns are remarkably consistent with both our theoretical mechanism and empirical evidence from Mexico: capacity-constrained retailers survive not by scaling up but by narrowing their scope and doubling down on their strengths.

2. See USDA and Iowa Public Radio.

3. Clearly, differentiation is not a viable strategy for local bookstores, as Amazon stocks virtually every title. On the other hand, specialization is.

Our findings speak to two influential theories. First, we revisit Waldfogel (2007)'s "tyranny of the majority," which argues that fixed costs cause markets to disproportionately serve majority tastes. We show that defensive specialization inverts this logic: as local generalists exit, surviving small stores retreat into niches, thereby improving the provision of niche products. Nevertheless, consistent with Waldfogel, we show in a model extension that specialization remains insufficient from a consumer welfare standpoint: consumers would be strictly better off if more (and larger) local stores specialized.

Relatedly, we qualify Anderson (2004)'s "long tail" theory. Anderson posits that capacity-unconstrained giants profitably operate in the long tail *because*, unlike physical stores, they face virtually no capacity constraints. We show that, in a world increasingly dominated by large retailers, the converse also holds: smaller retailers are *forced* into the long tail *precisely because of* their capacity constraints.

Our Contribution. To sum up, we make three contributions. First, we theoretically demonstrate that defensive specialization is optimal when – and only when – the capacity disadvantage is sufficiently large, and further study how it interacts with other strategic levers available to smaller retailers. Second, empirically, we provide causal evidence using empirical strategies tailored to the data availability: an IV strategy exploiting chain-level regional expansion costs and within-firm panel data with LLM-based specialization measures. Moreover, we use entrant size (supermarkets vs. convenience) and incumbent capacity to directly test our theoretical predictions. Third, we draw a host of implications for both retailers and consumers, complementing classic work on market structure and product variety.

The rest of the paper is structured as follows: Section 2 reviews the literature; Section 3 presents our theoretical results (which we extend in Appendix A); Section 4 describes the datasets; Sections 5 and 6 present the empirical evidence on market-level and within-store specialization. Section 7 concludes.

2. Related Literature

Our paper contributes to three related streams of literature. First, and foremost, the work on the impact of entry by large chain stores, including Igami (2011), Atkin, Faber, and Gonzalez-Navarro (2018), Ellickson, Grieco, and Khvastunov (2020), Arcidiacono et al. (2020), Lim et al. (2021), Talamas Marcos (2024), Caoui, Hollenbeck, and Osborne (2024), and Martínez-de Albéniz, Aparicio, and Balsach (2025). This literature documents how the expansion of productive chains reshapes local markets through exit, reallocation, and changes in format shares. More broadly, Bronnenberg and Ellickson (2015) and Hortaçsu and Syverson (2015) describe the longer-run shift toward larger, more efficient formats.

We extend this work by studying the impact of entry on local stores *as a function of their degree of specialization.* In doing so, we explain an empirical pattern in markets such as food retail in Mexico, where specialty retailers have *gained* market share in the face of chain entry – unlike generalist stores (Figure 3). Consistent with our analysis, Igami (2011) finds that in Tokyo the rise of supermarkets displaces mid-sized retailers, not the smallest ones. We contribute by showing that niche specialization – often infeasible or unprofitable for mid-sized stores – is a key reason why the smallest stores survive, and why such patterns disappear where specialization is not viable.

In particular, our empirical analysis has significant contributions beyond Talamas Marcos (2024), which studies the effect of the expansion of convenience chains in Mexico on neighborhood shops (general stores). First, we address a different question: while that work focuses on general stores that use differentiation as a response to competition (e.g., cultivating relationships with neighbors and offering informal credit), we examine the optimality of a defensive specialization strategy. Therefore, our analysis also incorporates the effects on specialty stores to assess whether the traditional sector becomes more specialized in response to competition from large entrants. Second, to test whether defensive specialization returns increase with the entrant's size (as our model predicts), we also estimate the effects of supermarket chain entries. And third, a crucial limitation of the data used in that work is that it does not allow observation of individual stores' product offerings. This paper addresses this data limitation by collecting two waves of surveys of retailers in Mexico City, allowing us to evidence within-store assortment specialization in response to competition from chains.

Our work also relates to research on product assortments, which is central to marketing strategy. This literature has long recognized that "consumers rank variety of assortment right behind location and price when naming reasons why they patronize their favorite stores" (Hoch, Bradlow, and Wansink, 1999). More specifically, our work speaks to the role of competition on equilibrium product assortments (e.g., Datta and Sudhir, 2011; Allcott et al., 2019). Hong, Misra, and Vilcassim (2016) highlight the interdependence of assortment decisions across categories, cautioning against narrow category-by-category analysis. We emphasize that *specialization* and *differentiation* are distinct strategic margins. Because the chain stocks every product in our baseline setting, differentiation is impossible; specialization – not differentiation – is the strategic lever. This modeling choice isolates the intensive/extensive margin forces that make specialization an equilibrium response to a large, full-line rival.

A related strand studies online–offline competition, which is typically characterized by large size asymmetries. Work on digital retail emphasizes the rise of assortment depth among capacity-unconstrained platforms (e.g., Anderson, 2006, Brynjolfsson, Hu, and Smith, 2010). Much of this literature examines how online entry shifts market share away from offline generalists. Our contribution is to show a complementary mechanism within offline markets: the entry of a large, full-line *offline* competitor can push small capacity-constrained stores into *offline* long-tail niches. In contrast to online markets – where the long tail arises from the *absence* of capacity constraints – we show that in offline markets it can arise precisely *because* of them.

Last, our findings speak to the literature on small firms in developing economies: the long tail of micro-enterprises (Hsieh and Olken, 2014), microenterprise survival (McKenzie and Paffhausen, 2019), and barriers to firm growth (McKenzie and Woodruff, 2017; Atkin et al., 2017; de Mel, McKenzie, and Woodruff, 2008). This literature typically interprets the persistence of small firms as driven by misallocation or binding constraints. Our mechanism offers a complementary explanation: some small retailers persist because defensive specialization provides effective insurance against chain entry. This helps explain why, as chains expand, generalists decline while specialists rise – ultimately leading specialty stores to surpass general stores as the primary food-purchasing destination for Mexican consumers (Figure 3). Defensive specialization can also inform policy and business-support programs such as those reviewed in McKenzie (2020).

3. Theory

Consider an economy with two stores, a (large chain) and b (local store); and z products divided into two different product categories, x and y. Store a carries all z products, while store b only has capacity for $k < z/2$ products. Although we assume that there exists only one local store, our intent is to model this as a typical local store, assuming that its effective competitor is the chain store. In Appendix A, we also consider the possibility of competition between local stores.

There is a measure one of consumers. Independently of preferences for specific products, consumers have a preference $\widetilde{w}$ for the local store b, reflecting tastes for "buying local," amenities, physical proximity, or aversion to chains. We assume $\widetilde{w}$ is uniformly distributed in $[0, w]$. The assumption of a 0 lower bound is without loss of generality: since the chain has a size advantage ($z/2 > k$), a positive $\widetilde{w}$ is required to buy from a local store, so consumers with $\widetilde{w} \leq 0$ always purchase from the chain.

To best illustrate the role of product assortment in our model, we start by assuming that product prices are constant and exogenously given, and with no further loss of generality we assume prices are equal to $1.

Consumers are equally split into two types, x consumers and y consumers.[4] x consumers (resp. y) like goods from product category x (resp. y) but derive no utility from goods from product category y (resp. x).[5] Specifically, in addition to the store preference utility $\widetilde{w}$, consumers derive utility $m(t)$ from purchasing at a given store, where t is the number of varieties of the consumer's preferred category. For example, if store b only stocks goods from product category x, then x consumers derive utility $m(k)$ from buying from that store, whereas y consumers derive utility $m(0) = 0$. Throughout the paper, we make the following standard assumption regarding $m(t)$:[6]

Assumption 1. $m(t)$ *is increasing and concave.*

One justification – but certainly not the only one – for Assumption 1: consumers have value $\widetilde{v}$ for goods in their preferred category, drawn from cdf $F(\widetilde{v})$ with support $[0, v]$. At a store with t products, a consumer buys the product yielding highest $\widetilde{v}$, receiving expected value $m(t)$ equal to the expected maximum of t draws from F. Standard results on order statistics imply $m(t)$ is increasing and concave (David and Nagaraja, 2003).

General vs. Specialty Store

The focus of our analysis is the local store's product strategy. We first consider the case when b pays no fixed cost to remain active. Should the store become a specialty store (stocking k products of one category) or a general store (stocking $k/2$ products of each category)? The first strategy sacrifices appeal for half of potential consumers; the second appeals to both types but at the expense of variety within each category.

Given our specification, profits for a general and a specialty store are respectively given by

$$\begin{aligned} \pi_g(z,k) &= \left(1 - \left(\frac{m(z/2) - m(k/2)}{w}\right)\right), \\ \pi_s(z,k) &= \tfrac{1}{2}\left(1 - \left(\frac{m(z/2) - m(k)}{w}\right)\right). \end{aligned} \tag{1}$$

The specialty-store strategy gives up half of all potential consumers but allows the local store to be more competitive in the chosen category (k products vs. the chain's $z/2$, rather than

4. In Appendix A, we consider the asymmetric case, that is, the case when one of the product categories has greater demand.

5. In Appendix A, we extend this to the case of eclectic consumers, who have positive valuations for both product categories.

6. Considering the large number of different variables used in the paper, Table 1 lists the main notation used in the paper.

Table 1
Main notation used in the paper

Variable	Description
a, b	large chain (e.g. Walmart) and local (small) stores
k, c	store b's capacity and cost per unit of capacity
$\widetilde{d}, d$	horizontal distance from local store b_0; $d = \max \widetilde{d}$
f, F	pdf and cdf of $\widetilde{v}$
g, s	general and specialty store
$m(t)$	maximum $\widetilde{v}$ from t draws out of $F(\widetilde{v})$
p	price
q	local's store market share
t	number of products
$\widetilde{v}, \widetilde{w}$	vertical and horizontal preferences (maximum values: v and w)
x, y	popular and niche product category
z	total number of products (carried by store a)
α, β	popularity of x, fraction of b's capacity devoted to x
π	store b's profit
τ	transportation cost (when b_0 and b_1 compete)

$k/2$ vs. $z/2$).

Before proceeding with the analysis, we make the following technical assumptions, ensuring interior solutions and non-trivial parameter spaces.

Assumption 2 (Value from broad variety) $v > 2m(k) - m(k/2)$.

One way to interpret Assumption 2 is that v – the maximum possible value from accessing the full product space – needs to be sufficiently high. Absent this, the chain's variety advantage would be too weak to generate the competitive pressure that drives specialization.

Assumption 3 (Interior solution) $2(m(k) - m(k/2)) < w < v - m(k/2)$

Assumption 3 ensures that the solution is interior. The lower bound ensures that a general-store strategy can be profitable when the chain is small. The upper bound ensures that generalist profits reach zero at some finite chain size. Under Assumption 2, this interval is non-empty. When the upper bound fails to hold, we are in a corner solution whereby it is a dominant strategy for b to be a general store. If Assumptions 2–3 hold, however, then the choice of general or specialty store depends on the relative values of z and k, as stated in the following result:

Proposition 1. *There exists a threshold $z_{gs} = z_{gs}(k, w)$ such that an active firm b optimally chooses to be a specialty store if and only if $z > z_{gs}$. Moreover, $z_{gs}(k, w)$ is increasing in both k and w.*

Proof: The proof for this and all other theoretical results can be found in Online Appendix A. ■

Proposition 1 can be equivalently stated in terms of local store heterogeneity, as shown in the following corollary:

Corollary 1. *There exists a threshold $k_{gs} = k_{gs}(z, w)$ such that an active firm b optimally chooses to be a specialty store if and only if $k < k_{gs}$. Moreover, $k_{gs}(z, w)$ is increasing in z and decreasing in w.*

Proposition 1 and Corollary 1 offer two readings of the same result: a local store optimally specializes if and only if its size disadvantage relative to the chain entrant is sufficiently large. This relationship is moderated by consumer preferences for store b: the larger w, the greater the size disadvantage must be to justify specialization.

The intuition involves an "extensive margin" vs. "intensive margin" tradeoff. By specializing, a store forgoes half its potential customers (extensive margin) but increases expected quality for remaining customers by stocking twice as many items in one category (intensive margin). As z increases, the chain becomes more attractive, hurting general stores more than specialty stores – the general store loses customers from both categories, whereas the specialty store only loses from a smaller set. Starting from indifference, there exists a threshold past which specialty dominates.

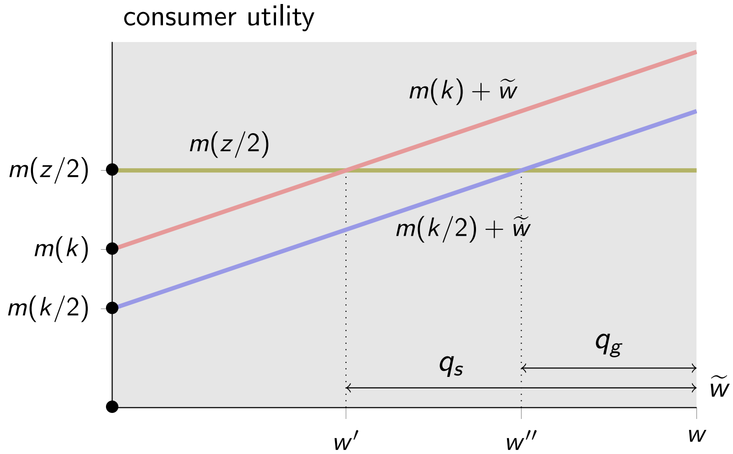
Figure 1 illustrates the tradeoff. The horizontal axis measures $\widetilde{w}$ (higher = stronger preference for local store); the vertical axis measures store value. The four bullet points show utility from: a specialty store in the wrong category (zero); a general store ($m(k/2)$); a specialty store in the right category ($m(k)$); and the chain ($m(z/2)$). Thresholds w' and w'' mark where consumers switch to the specialty and general store, respectively.

The figure shows the indifference point where $q_s = 2q_g$ (specialty captures twice the share of its potential customers as general does). As z increases, both q_s and q_g decline at the same absolute rate, but since $q_g < q_s$, this represents a larger proportional drop for the general store. Thus, past indifference, specialty dominates.

Survival

We now extend our analysis to allow for the possibility of exit. Suppose that the local store must pay a fixed cost $c\,k$ in order to operate, where c is cost per unit of capacity. The

Figure 1
Choice of general vs specialty store



cost parameter c is drawn from a continuous distribution $G(c)$ with positive support that is independent of store type (general vs. specialty). Store profit, conditional on the realized cost parameter c, is then given by

$$\begin{aligned} \pi_g(z,k) &= \left(1 - \left(\frac{m(z/2) - m(k/2)}{w}\right)\right) - c\,k \\ \pi_s(z,k) &= \tfrac{1}{2}\left(1 - \left(\frac{m(z/2) - m(k)}{w}\right)\right) - c\,k \end{aligned} \tag{2}$$

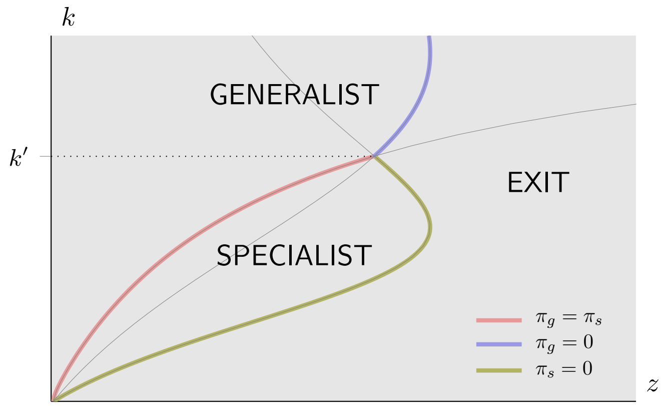
In this setting, the local store has three choices: remaining open as a general store, remaining open as a specialty store, and exiting. In order for the store's choice not to be trivial, we assume that a switch from general to specialty (or from specialty to general) implies a strictly positive cost (which however can be arbitrarily small). Since each store draws its cost parameter c independently, stores with higher realizations of c are more likely to exit.

Proposition 2. *There exists a threshold $\bar{z}$ such that, for $z > \bar{z}$, upon entry by a large store, general stores are more likely to exit than specialty stores.*

Figure 2 illustrates the equilibrium with exit (assuming $\widetilde{v}$ uniform). The red line is the indifference condition $\pi_g = \pi_s$; the blue and green lines are zero-profit conditions for general and specialty stores. The three lines cross at the same point (when $\pi_g = \pi_s = 0$). This yields three regions: GENERAL (NW), SPECIALTY (S), and EXIT (E).

For $k < k'$ (where k' is indicated in Figure 2), as chain size increases, firm b optimally switches from general to specialty store; with further increases, it eventually exits. Combined, Propositions 1 and 2 imply:

Figure 2
Comparative statics with respect to z and k when exit is a possibility



Corollary 2. *Upon entry by a large chain, small enough general stores either exit or become specialty stores.*

The model developed so far is intentionally parsimonious: it allows us to identify the main effects at work and to derive testable empirical implications. Specifically, our theory makes three testable predictions regarding the effects of entry by a large chain into a local market:

- Reduce the local stores' profitability, particularly that of local general stores
- Lead local stores to become more specialized, especially *smaller* stores or those facing the entry of a *larger chain* competitor
- Increase the exit probability of local stores, especially general stores

Nevertheless, it is both interesting and important to understand the role played by our assumptions, and to relax several of them by endogenizing more small store decisions and considering different consumer types, as well as richer forms of competition. In particular, as mentioned earlier, our basic model is based on a series of simplifying modeling choices: so far, we have considered competition between *one* small store and *one* large entrant; we assume away price competition to zoom in on stocking decision; we assume each individual consumer has a preference for *exactly one* of two (symmetric) product categories; and we assume an *exogenously given* preference for small stores.

To respect the journal's page limit, the rest of our theoretical analysis can be found in Appendix A. There, we prove the robustness of our model to a host of alternative assumptions. Extending each of the above, we study *i*) asymmetric product categories; *ii*) endogenous prices; *iii*) eclectic consumers, who value both product categories; *iv*) competition between two symmetric stores (as well as a larger chain entrant); and v) endogenous investments in amenities.

In short, here is a subset of our findings: small retailers complement specialization with higher prices (compared to generalists), capturing an even smaller market share than they otherwise would, but increasing profits (Proposition 4; we label this "boutique effects"); eclectic consumers – consumers who are interested in more than one product category – can either hurt or help small stores; when stores can invest in amenities to differentiate from large chains,[7] i) they invest more when size disadvantages are larger and ii) for a given size disadvantage, generalists invest more than specialists (Proposition 5); competing with both local stores and large chains strengthens specialization incentives (Proposition 6). Last, in the spirit of Waldfogel's tyranny of the majority, we show that, when the small store is indifferent between being a general or a specialty store, the average consumer strictly prefers the latter (Proposition 7).

Next, we turn to describing our data and to testing our main theoretical predictions and derive additional results of interest in the context of entry by large chains in local markets in Mexico.

4. Data

Background. There are two food retail channels in Mexico: the traditional and the modern. The traditional sector is characterized by a large number of small, often family-owned and operated businesses. With over one million establishments engaged in the retail sale of groceries, food, and beverages, and over 500,000 outlets that sell prepared food, the traditional sector plays a crucial role in the economy. These establishments represent 34% of all establishments, 15% of employment, and 4% of value added in the country (Economic Census, 2019).

Within the traditional sector, general and specialty stores coexist. While similar in size, general small stores, often referred to as *neighborhood shops*, sell a wide variety of product categories, for example, soft drinks, milk, cheese, ham, snacks, cigarettes, bread, and household and kitchen products (see top left of Figure B.1 for a visual example). The general small shop is the most common establishment in the country. There are nearly 600,000 of these, and because of their large number, they have almost 30% market share of the food and beverage retail industry. However, this market share has declined by 10% in the last two decades (see Figure 3).

The other component of the traditional sector is the small specialty stores. These stores

7. See for instance Raffaelli (2020), Kravitz Hoeffner (2022) and Maurice-Jones (2025).

are also prevalent, with over 400,000 establishments nationwide. Unlike general small stores that sell multiple product categories, specialty stores are characterized by offering mainly one product category: bakeries, tortillerias, butchers, poultry shops, seafood retailers, dairy stores, and fresh produce stores, among others (see Figure B.2 for visual examples). Over the past 20 years, specialty stores have increased their market share in the food sector by more than 30%, establishing themselves as the leading choice for food purchases nationwide (see Figure 3).

While similar in size, small specialty and general stores differ starkly in their product offerings: general stores offer more product categories but fewer varieties within each, while specialty stores focus on one product category and offer more variety. Table B.1 compares general and specialty stores of the traditional sector in 1999, before most convenience and supermarket chain entries. General and specialty stores were quite similar in size. There was no difference in revenue, and slight differences in profits and age. Specialty stores had 5% lower profits and employed 2% fewer workers, with no difference in revenue per worker. Yet, these two types of establishments differ in that some are generalists with multiple product categories, and the others are specialists focusing on one product category.

On the other hand, the modern channel in Mexico encompasses supermarkets and convenience stores. These two store formats are larger than those in the traditional sector, and both are generalists, offering multiple food product categories. Convenience and Supermarket chains have considerably expanded over the past two decades. The two most prominent chains, Oxxo and Walmart, have established over 20,000 and 3,000 retail outlets, respectively. This proliferation of chains has significantly altered the retail landscape in Mexico (see Figure B.3).

Regarding the physical size of food retailers in Mexico, traditional sector establishments are the smallest, convenience store chains are of medium size, and supermarket chains are the largest. Traditional sector stores are approximately 30 square meters in size, employ fewer than two people (mostly owners and family members), and are primarily located next to the owners' homes (Economic Census, 2019). The most common convenience chain stores are Oxxo, 7-Eleven, Circle K, and 3B, whose retail space is about 100 square meters, but reaches around 400 square meters when storage and parking are included, and employ, on average, 5.7 people per store (Grupo Reforma (2022); Economic Census (2019)). Supermarkets are the biggest food retailers, with an average retail space of over 2,500 square meters and more than 70 workers per store (ANTAD (2022), Economic Census (2019)).

Data Overview. The empirical analysis relies on two waves of detailed surveys of retail stores in Mexico City collected by the authors, nationwide microdata collected and sheltered by the Mexican Statistics Institute (INEGI), and public data. We accessed INEGI's

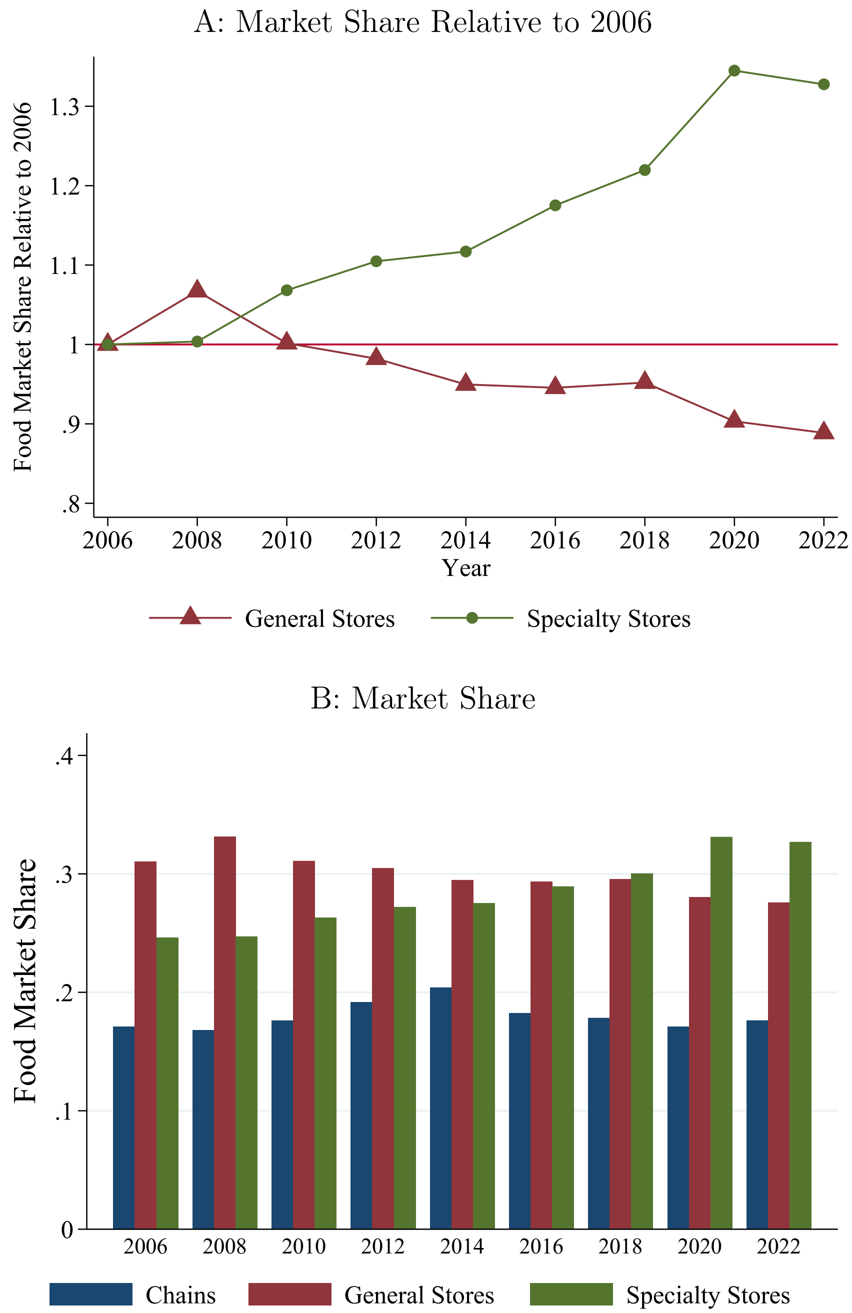


Figure 3
Market Share by Store Format in Mexico's Food Market

Source: National Income and Expenditure (ENIGH) Surveys (2006 - 2022)
Note: Panel A displays food expenditure shares by store format relative to their 2006 share. Panel B presents shares by store type. All food expenditures are included (category "A" in the data). Chains include supermarkets, membership clubs, department stores, and convenience stores. General stores include neighborhood shops, and specialty stores are establishments that focus on a single product category (as defined by ENIGH). The calculations use ENIGH's representative sampling weights.

microdata through the institution's secure microdata lab in Mexico City.

Nationwide Microdata. The microdata from INEGI are the Economic Censuses (1999, 2004, 2009, 2014, 2019) and the Income and Expenditure Surveys (2008, 2014, 2018). The Economic Censuses cover all physical establishments in the country–both formal and informal, of any size. During the economic census data collection, enumerators classify each establishment based on its product offerings into one category of the North American Industrial Classification System for Mexico (SCIAN), which is very similar to the one used in the U.S. (NAICS). This detailed classification of food retailers allows us to distinguish between general and specialty stores.

Using the 6-digit SCIAN classification we identify specialty establishments as those that specialize on one product category: red meat (461121), poultry meat (461122), fish and seafood (461123), fresh fruits and vegetables (461130), edible seeds and grains, spices, and dried chiles (461140), milk and other dairy products (461150), sweets and pastries (461160), popsicles and ice cream (461170), other food products including bakeries, snack bars, and *tortillerias* (461190), wine and spirits (461211), beer (461212), non-alcoholic beverages and ice (461213), and cigarettes, cigars, and tobacco (461220).[8,9] To ensure we include only establishments in the traditional sector, we exclude those owned or operated by the government or a chain.[10]

Similarly, we identify general stores as those that offer a wide variety of food categories based on their 6-digit SCIAN classification. In the modern sector, which includes chains, there are supermarkets (462111) and convenience stores (462112), while in the traditional sector, there are neighborhood shops (461110). For example, the code 461110, used for neighborhood shops, is for "grocery, variety, and miscellaneous stores primarily engaged in retail trade of a wide variety of products such as milk, cheese, cream, cold cuts, candies, cookies, bread, pastries, snacks, fried foods, canned goods, bottled purified water, soft drinks, beer, bottled wines and liquors, cigarettes, eggs, toilet paper, detergent, soap, paper napkins, and disposable kitchen utensils" (INEGI, 2025).

Figure B.1 in the Online Appendix shows examples of each of these generalists, where the key difference between them, size, becomes apparent. On average, convenience stores employ more than three times as many people and generate more than 20 times the revenue of neighborhood shops; yet, supermarkets employ more than 40 times as many people and

8. See Figure B.2 for visual examples.
9. For robustness, we use two alternative classifications of specialty stores: i) establishments within retail sale of groceries and food (4-digit classification 4611), which excludes alcoholic drinks and tobacco, and ii) establishments within beverages, ice, and tobacco (4-digit classification 4612). The results are robust to these alternative classifications.
10. However, our results are robust to including specialized chains in our estimation.

create more than 350 times the revenue of neighborhood shops Economic Census (2019). We focus on the entry of supermarkets and convenience stores that belong to a chain, because for those, our instrument can leverage a cost shifter at the chain level. We define a supermarket chain as one with more than 20 establishments at any point in time, and a convenience chain as one with over 100 establishments.[11]

Starting in 2009, INEGI added an establishment identifier to the Economic Censuses. To track establishments before 2009, we use the establishment identifiers created by Busso, Fentanes, and Levy (2018). We use this establishment-level panel from 1999 to 2019 to identify entry and exit.

The Income and Expenditure Surveys (ENIGH) contain data on what households buy and in what type of establishment they buy it. While the ENIGH is publicly available, we access it through INEGI microdata lab to gain access to information of where the household lives down to the census tract (AGEB) level. We identify transactions in specialty stores as those where the purchase location, categorized by INEGI, is a specialty store, defined as "[e]stablishments dedicated to the commercialization of a single line of business – that is, they sell one type of product or service" (ENIGH, 2006). The sample of the ENIGH has grown throughout the years. In 2006, it contained responses from a little more than 20,000 households; by 2018 it included more than 70,000 responses.

INEGI's geo-statistical framework for urban Mexico divides the country into states, municipalities, localities, and urban census tracts (AGEBs). The total number of AGEBs ranges from 37,000 to 47,000 (depending on the census year). Each AGEB typically consists of 25 to 50 blocks, which encompass an average of 650 households and around 2,000 residents. AGEBs are precisely defined by streets, avenues, or other easily identifiable characteristics in the field. This design by INEGI is intended to facilitate the data collection process conducted by enumerators.

For the analysis using Economic Census data, the market definition in our main specification is the census tract. However, for the analysis based on the ENIGH data, we need to construct larger markets, as few census tracts are surveyed across multiple waves of the ENIGH. Without observing each market multiple times, we cannot include a market fixed effect in the estimation. Therefore, we use census tracts to define markets by drawing a buffer (circle) around each tract's centroid, and a market is the union of tracts that overlap with that buffer.[12] We draw these buffers with a radius of 1 km for our main specification

11. We add the size restriction in the definition of chains to increase the likelihood that chains operate in multiple cities and that they are likely to benefit from regional economies of scale, which is the cost shifter used in the instrument.

12. We use the union of census tracts instead of only the buffer because the census tract is the smallest measure at which we observe the household location.

and 1.5 and 2 km for robustness.

Primary Data Collection. We conducted two waves of in-field business surveys among 554 local retail stores in Mexico City in 2018 and 2020, prior to the onset of the COVID-19 pandemic. Our sampling strategy focused exclusively on local stores (i.e., no chains) with up to 10 employees. Our field team canvassed all major commercial areas of the city to invite any store fitting this profile to participate in our study. The resulting sample consists of those who consented to participate in this multi-year data collection exercise, and we discuss the representativeness of this sample further in this section.

During the baseline survey, we collected data on: (i) the location of the store through GPS coordinates; (ii) the type of store based on the North American Industrial Classification System (SCIAN) code and an open-text description of the primary products and categories sold; (iii) the total number of SKUs; and (iv) the top 3 SKUs sold at the store by volume. During the endline survey in early 2020, we asked store managers whether they had added any new products to their assortment and whether they had removed any existing products from their assortment. If they reported either, we collected text data on which products they removed or added. Additionally, we asked managers to report whether the *primary* reason for any assortment change was competition, demand, marginal cost or margin considerations, or any other reason (e.g., a supplier issue). In Online Appendix B, we report the phrasing of these survey questions and the structure of responses obtained. For open-text responses, the enumerator transcribed the owner's response, noting concrete evidence or examples. In Table B.2, we display some exemplar text responses on product additions and removals from five randomly-sampled food retailers.

Data quality was maintained through both training and supervision. Enumerators were instructed to obtain visual proof of product additions before recording them in text and to probe for specific evidence when products were reported as removed. Team leaders and a research manager oversaw fieldwork. At the end of each field day, team leaders reviewed all completed surveys for inconsistencies or entry errors in the electronic tools. Because teams were still located near the businesses they had audited, any issues could be corrected through immediate revisits.

Figure B.4 maps the surveyed stores in Mexico City, and Table B.3 provides descriptives on store characteristics. The geographical spread of the stores surveyed, pictured in Figure B.4, allows us to obtain useful variation in nearby chain entry and ensures better representativeness of the city's retail landscape. Also, Table B.3 shows that our sample closely aligns with citywide means from the 2019 Economic Census for monthly revenues across the main retail sub-sectors represented (461 groceries, 465 stationery and office supplies, and 722 prepared food). Surveyed firms employ slightly more workers on average than census establishments. We included retailers outside the food sub-sectors who still compete with supermarkets to achieve a larger sample size, but all our results hold when restricting the analyses to the food-only sample.

We merge this panel dataset from the field with the microdata from INEGI's firm registries (DENUE) to identify entries of chain competitors – both supermarket and convenience chains – between baseline and endline. A chain is considered a competitor if it enters within pre-defined radii (500 m, 1 km, and 2 km) that correspond to trading areas in urban environments where stores primarily rely on walking traffic.

Public Data. For one of our identification strategies, we use street width data from Open Street Maps. We classify trunk, primary, secondary, and tertiary streets as *wide* and the remaining categories as *narrow*. In the resulting classification, 21% of the total street length is *wide*. The remaining 79% of streets are *narrow*, and the vast majority of the narrow streets, 95%, are residential streets. We construct a measure of the prevalence of wide streets by adding the lengths of all wide streets in the market and dividing it by the market size, specifically dividing by the square root of its area.

5. Market-level evidence of specialization

Section 5 and Section 6 contain our empirical analysis testing Propositions 1 and 2. This section provides evidence at the market level using nationwide microdata, and the following section includes the establishment-level findings using the two waves of detailed surveys. We start by introducing our empirical strategy and estimating equations. Then, we transition to the empirical evidence. We show that the entry of a chain (whether a supermarket or a convenience store) negatively affects the average revenue and profits of traditional general and specialty retailers. The effects are larger for supermarket entries and increase with the entrant's proximity. Consistent with our theoretical predictions, the effects are significantly larger for general than for specialty stores. Then, we show that as a result of the entry, the traditional sector becomes more specialized. The number of general stores decreases, the number of specialty stores increases, and the share of specialty-store revenue increases. Consistent with our theoretical prediction, when the entrant is larger (higher z, i.e., a supermarket), we observe a bigger shift towards specialization at the market level.

5.1. Empirical strategy

Chains are typically drawn to markets with high demand and foot traffic. These are also the markets where shops in the traditional sector are more likely to be successful, leading to an endogeneity problem in estimating the relationship between the number of chain stores in a market and outcomes for traditional stores. Including market fixed effects can alleviate this concern by controlling for market-specific time-invariant characteristics. However, our nationwide data spans 20 years, and we observe firms only every 5 years, potentially leaving the estimation vulnerable to bias from market-specific shocks that affect the profitability of chains and traditional sector stores within these 5-year windows.

To alleviate this concern, we follow an identification strategy based on instrumental variables introduced in Talamas Marcos (2024) that builds on the intuition of Jia (2008) and Holmes (2011), where stores of the same chain in the area share costs (e.g., logistics, marketing, and overhead). Therefore, the expected profitability of a potential new store increases with the addition of stores from the same chain in nearby cities, as the average cost declines with the additional store sharing costs.

The instrument is based on the interaction between two components. The first one captures the regional economies of scale (cost sharing) of chains, and it is the first term of the product in equation 3. Specifically, it is the square root of the sum of the squared lagged number of stores per chain in nearby municipalities (a Herfindahl-Hirschman Index without normalization), where nearby municipalities are the second-degree neighbors (adjacent municipalities and those adjacent to them).[13] This measure increases with the number of same-chain stores in nearby cities, and it provides variation at the municipality and year level, but does not predict where new chain stores will locate within municipalities.

The second component incorporates variation within cities based on the suitability of the market for chains based on the prevalence of wide streets. While establishments in the traditional sector are often next to the owner's house, chains enter on wide streets and at intersections to target traffic. We measure the suitability of a market for chain stores using its total length of wide streets divided by the square root of the market area (second term in the product in equation 3). The instrument we use is the product of these two components,

$$Z_{m,t} = \underbrace{\left(\sum\nolimits_f (\#StoresNearbyMun_{f,c,t-1})^2\right)^{1/2}}_{\text{Regional Economies of Scale}_{c,t}} \times \underbrace{\frac{Total\ wide\ streets\ length_m}{Area_m^{1/2}}}_{\text{Prevalence of Wide Streets}_m}, \tag{3}$$

where Z is the instrument, m denotes the market which belongs to municipality (city) c, t denotes the census year, and f denotes the firm.

Our first and second stages estimating equations are

13. In the robustness section we show that the results are robust to using third-degree neighbors as well.

$$CS_{m,t} = \gamma_1 Z_{m,t} + \zeta_m + \eta_{c(m),t} + \mu_{m,t}. \tag{4}$$

$$Y_{m,t} = \beta_1 \widehat{CS}_{m,t} + \zeta_m + \eta_{c(m),t} + \epsilon_{m,t}, \tag{5}$$

where $Y_{m,t}$ is the outcome of interest (e.g., number of general establishments, number of specialty establishments, share of specialty establishments, and specialty stores' share of revenue). CS stands for the number of chain stores. The estimation includes market fixed effects, ζ_m, and municipality-year fixed effects, $\eta_{c(m),t}$. Standard errors are clustered at the municipality level because the regional chain expansion varies across municipalities.

We estimate the effect of the entry of a convenience chain and a supermarket chain separately. Therefore, we construct an instrument for the number of convenience chains in the market and another one for the number of supermarket chains. Estimating the effects separately enables us to assess whether the entry of a larger general store has a greater impact on the degree of specialization, as predicted by Proposition 1.

First Stage. The instrument is a strong shifter for the number of chain stores in the market. For convenience chains, a one-standard-deviation increase in the instrument is associated with an increase of 0.1 (33%) convenience chain stores in the census tract. For supermarket chains, a one-standard-deviation increase in the instrument is related to an increase of 0.02 (20%) supermarkets in the census tract. Table C.1 displays the first stage estimates. Throughout the results section, the tables display the Kleibergen-Paap F-statistic, which is large, ranging from 40 to over 100, indicating that the first-stage of the IV is strong.

Exclusion Restriction and Potential Identification Concerns. The instruments' exclusion restriction is that traditional general and specialty stores in areas favorable for chains are affected by an increase in the number of chain stores in adjacent municipalities *only* due to the heightened likelihood of a chain establishing itself in their neighborhood. Since the cost advantages of nearby chain stores are specific to the chain that opens them, the exclusion restriction is plausibly satisfied. However, there are several potential concerns.

A potential first concern is that chains are entering cities that are growing faster, which will also be the cities where traditional-sector sales are increasing. City-year fixed effects control for this and other city-level potential shocks. A second concern is that chains enter affluent neighborhoods, where traditional sector shops are often more successful. Or, similarly, neighborhoods with wide streets may be more affluent, and it's there that chains would prefer to enter. Market fixed effects address this issue and others related to cross-sectional time-invariant differences across markets.

A third possible concern is that chains may enter cities that are growing faster in markets where wide streets are more prevalent, which may also be the markets where income and demand are growing faster. This is unlikely to be an issue because while the entry of same-chain stores in nearby cities predicts openings in a city, competitors' openings in nearby cities do not.[14] Therefore, the entry into a city is likely driven by chain-specific characteristics rather than city-wide ones. Additionally, our main outcomes of interest are in relative terms, e.g., the share of specialty stores in the market. Thus, if something affects general and specialty stores in the traditional sector in a similar manner, it will cancel out.

A fourth possible concern is that the exclusion restriction may be violated if an additional store in a neighboring city reduces the costs of distributing to traditional-sector stores of large suppliers (e.g., Coca-Cola), and the supplier passes along the savings to those stores. This is unlikely to be a significant concern because any savings and pass-through would be divided among all retailers, and it would be minimal at the store level. Additionally, if it affects all food retailers in the nearby cities, it would cancel out with the city-year fixed effects.

5.2. Results

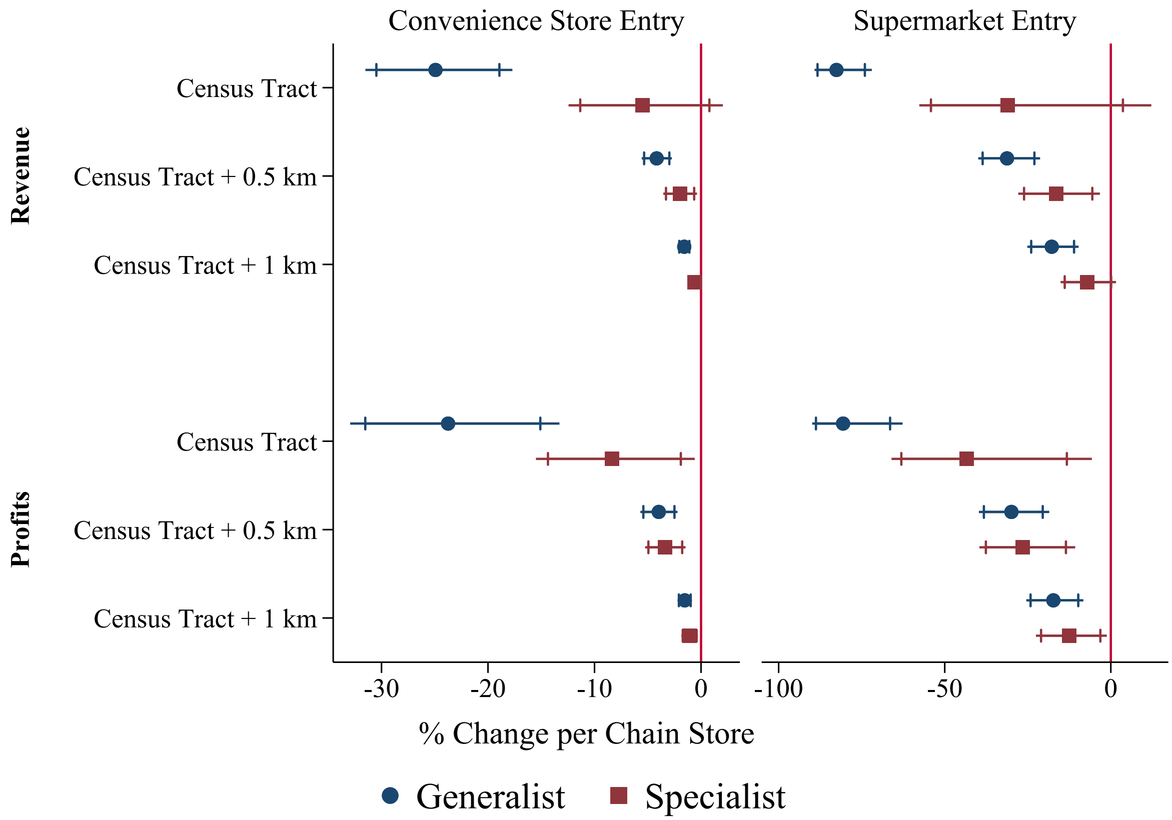
Average Profits and Revenue. The entry of convenience store and supermarket chains lowers the average profits and revenues of food retailers in the traditional sector (Figure 4). The decline is significantly larger for general stores. An additional convenience chain store in the census tract entails a reduction in average profits and revenue of general stores of almost 30%, while specialty stores suffer a decline of close to 10%. The supermarket entry has a similar pattern, but the magnitude of the effect is much more prominent. General stores average profits and revenue decline by 80%, while specialty stores experience a decline between 40% and 50%. As expected, these effects shrink as the market size increases because the average distance to the entrant increases.

These findings align with the intuition behind proposition 1, which states that the chain impacts both general and specialty stores; however, it hurts the general store more because it loses customers across all categories, whereas the specialty stores only lose customers from a smaller segment. If the size disadvantage is significant enough ($z >> k$), the effect will be large enough that it will no longer be optimal to operate as a general store, but instead as a specialty store. We test this predicted transition to specialization at the market level next.

Market Specialization. The entry of a convenience chain store or a supermarket increases the degree of specialization in the traditional sector (Table 2). In particular, the percentage of specialty stores increases by 4.7 percentage points (pp; 17%) with the entry of a convenience chain store and 29 pp (104%) with the entry of a supermarket in the census tract.

14. See Online Appendix C of Talamas Marcos (2024) for details on this analysis.

Figure 4
Effect of the Entry of a Large Chain



Note: The figure displays the effect of the entry of a convenience chain store or a supermarket chain on the revenues and profits of traditional general and specialty stores, estimated using equation 5 using 2SLS. The estimates are for three market sizes: census tracts, census tracts within 0.5 km, and census tracts within 1 km. Standard errors are clustered at the municipality level. All coefficients are displayed with their 90% and 95% confidence interval.

The main driver of the effect is a steep decline in the number of general stores, which does not occur for the number of specialty stores. The number of general establishments declines by 2.5 with the entry of a convenience chain store and by 15 with that of a supermarket. On the other hand, the number of specialty stores does not decline; it even slightly increases. We discuss this increase in more detail later when analyzing the effects on entry and exit.

As shown in Figure 4, the entry of a supermarket has a larger effect on the traditional sector than the entry of a convenience chain. This is consistent with the idea that a larger entrant would steal more business from the incumbents. This does not necessarily imply that the supermarket would also have a larger effect on market-level specialization (i.e., the ratio of specialty to the sum of specialty and general). For example, if the percentage effect for both specialty and general stores were the same, the ratio of specialty over the total of specialty plus general would remain unchanged, implying a null impact on specialization. Therefore, our specialization effects capture the asymmetry in the effects on specialty and general stores.

Table 2
Entry of Large Chains and Specialization in the Traditional Channel

	Convenience Chain Entry				Supermarket Chain Entry			
Dependant Variable:	% Specialized (N)	Generalist Stores (N)	Specialized Stores (N)	% Spec. Revenue	% Specialized (N)	Generalist Stores (N)	Specialized Stores (N)	% Spec. Revenue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Chain Stores	4.73***	-2.49***	0.88*	5.72***	29.11***	-15.34***	5.43**	35.21***
	(0.73)	(0.38)	(0.45)	(0.86)	(5.79)	(2.12)	(2.67)	(6.28)
Observations	160,103	160,103	160,103	160,038	160,103	160,103	160,103	160,038
Year x Mun. FE	Y	Y	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y	Y	Y
Mean Dep. Var	27.9	12.8	8.0	32.1	27.9	12.8	8.0	32.1
Mean Ch. Stores	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0.1
KP-F Statistic	99.8	99.8	99.8	99.8	55.0	55.0	55.0	55.1

Note: The table displays the effect of the entry of a convenience chain store (columns 1-4) or a supermarket (columns 5-8) at the census tract level, estimated based on equation 5 using 2SLS. The % Specialty (N) is the number of specialty stores divided by the total general and specialty stores in the traditional sector in the census tract. The % Spec. Revenue is the total revenue of specialty stores divided by the total revenue of general and specialty stores in the traditional sector in the census tract. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

In our model, the entrant's size, z, matters – Proposition 1 shows that bigger entrants (larger size disadvantage) lead to a stronger incentive to specialize. This is indeed what we observe empirically when comparing the effect on specialization of the entry of a convenience chain store and a supermarket (Columns 1 and 5 of Table 2). The entry of a supermarket increases the share of specialty retail in the traditional sector six times more than the entry of a convenience chain store. Overall, the market becomes more specialized in a broad range of measures. Table 2 shows that the share of specialty establishments and the share of revenue in specialty establishments within the traditional sector increase. Similarly, the share of profits, value-added, employment, and hours worked of specialty establishments in the traditional sector increases (Table C.2).

Consistent with our estimates based on the Economic Censuses, estimates from household income and expenditure surveys indicate that the entry of a convenience chain store increases the share of consumption in specialized stores within the traditional retail sector. Table C.3 shows that each entry of a convenience chain store in the market (census tracts within 1 km) increases the share of consumption in specialty stores by 0.4 pp and the share of trips to specialty stores by 0.4 pp. These estimates are robust to alternative market definitions, such as all census tracts within 1.5 and 2 km, and as expected, the magnitude of the estimates declines with larger market sizes (Table C.4).

Ordinary Least Squares Estimates. Estimating the effect of a chain's entry on the number of traditional-sector retailers without fixed effects or an instrument would be subject to omitted-variable bias, e.g., higher demand would lead to more traditional-sector retailers and chain entries. However, our main results concern the degree of specialization in the market and, therefore, are relative. The effect on the share of specialized establishments encompasses changes in the numbers of general and specialty stores. If both of these estimated effects were upward-biased, this bias would, to some extent, cancel out in the measure of the share of specialized establishments. This is precisely what we observe when we estimate without fixed effects and an IV. The effects on the number of general and specialty stores are upward biased, yet the effect on the share of specialty stores in the census tract is consistent with that of our main specification (see Table C.8).

Specialization and Firm Dynamics. We now analyze the effect of a chain's entry on the degree of specialization through the lens of firm dynamics. We find that the entry of a chain results in an increased likelihood of traditional sector exits being by general stores and entries being by specialty stores. Table 3 shows that the entry of a chain leads to an increase in the probability of an entrant being a specialty store by 4.2 pp with an additional convenience chain store and by 26 pp with an additional supermarket. Exit also contributes to the market's specialization. The entry of a chain leads to a decrease in the probability of a store closure being a specialty store by 2.6 pp with an additional convenience chain store and by 13 pp with an additional supermarket.

These estimates aim to illustrate the effects through business dynamics. Still, they could also include the effects through within-firm specialization, because when a generalist pivots into a specialty store, it may be recorded as an exit and an entry in the data. Section 6 investigates and isolates the effect on within-firm specialization by leveraging our two waves of data collection of retail establishments in Mexico City.

Robustness. The finding of the entry of a large chain leading to an increase in specialization in the traditional sector is robust to alternative market specifications, instrument constructions, and classifications of specialized establishments. Table C.5 displays the estimates of the effect of the entry of a convenience chain store on the share of specialized establishments within the retail sector with different market sizes: census tract (baseline specification), census tracts within 0.5 km, 1 km, 1.5 km, and 2 km. All of these specifications lead to the same conclusion; however, as expected, the effect decreases as the definition of the market size increases because the average distance to the entrants increases. Table C.6 replicates Table 2 but using third-degree neighboring cities instead of second-degree neighboring cities to construct the instrument. The estimates are similar and consistent.

Table 3
Entry of Large Chains and Specialty Share of Entries and Exits

Dependant Variable: Specialty Share of Entries and Exits				
	Convenience Chain Entry		Supermarket Chain Entry	
	Entry	Exit	Entry	Exit
	(1)	(2)	(3)	(4)
Number of Chain Stores	4.25***	-2.62*	26.17***	-13.63*
	(1.12)	(1.47)	(6.78)	(7.56)
Observations	148,582	101,470	148,582	101,470
Year x Mun. FE	Y	Y	Y	Y
Market FE	Y	Y	Y	Y
Mean Dep. Var	35.4	32.7	35.4	32.7
Mean Ch. Stores	0.3	0.3	0.1	0.1
KP-F Statistic	101.9	88.9	49.3	60.2

Note: The table displays the effect of the entry of a convenience chain store (columns 1 and 2) or a supermarket (columns 3 and 4) at the census tract level, on the share of entry and exit of specialty stores in the traditional sector, estimated based on equation 5 using 2SLS. The specialty share of entries is the number of entries from specialty stores divided by the total number of entries from specialty stores and general stores in the census tract. The specialty share of exits is the equivalent measure for the exits. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.7 uses two alternative classifications of specialty stores: establishments within retail sale of groceries and food (4-digit classification 4611) and establishments within beverages, ice, and tobacco (4-digit classification 4612). The results are robust to these alternative classifications, with estimates slightly larger for the first and slightly smaller for the latter.

6. Within-firm evidence of specialization

We leverage our two waves of data collection on retail establishments in Mexico City to provide within-firm evidence that upon entry by a large chain, surviving local stores optimally shift towards specialization. We particularly focus on testing Corollary 1 of our theory, which implies that the shift toward specialization should: (i) be present in smaller, local stores that are more capacity-constrained (small k, per the model) and (ii) should not be present in larger, local stores. We first describe our empirical strategy. We then present empirical evidence demonstrating that local stores respond to chain entry with assortment changes explicitly driven by competition or customer demand, in line with our theory. Conversely, we document null effects for assortment changes reported for placebo reasons unrelated to our theoretical mechanisms (e.g., supplier issues or input cost changes). We further show that these assortment adjustments occur in response to competition in *only* small, capacity-constrained stores (below-median number of SKUs), consistent with the theoretical emphasis on relative capacity disadvantages. Finally, we introduce an LLM-based specialization index to explicitly demonstrate that assortment changes in response to large chain entry are in the direction of greater specialization.

6.1. Empirical strategy

Our primary dataset is collected entirely within Mexico City and therefore does not permit the instrumental variables strategy in Section 5, which relies on cross-municipality variation. We instead estimate first-difference specifications that exploit within-store changes in product assortments between baseline ($t = 1$) and endline ($t = 2$). This approach absorbs time-invariant store characteristics that may be correlated with chain entry. The short interval between survey waves further limits concerns about time-varying local shocks. The mean time elapsed between baseline and endline is 19.7 months (SD = 2.6), substantially shorter than the five-year gap between census waves. We trim the top 1 percent of the elapsed-time distribution to exclude larger delays, which primarily reflect difficulties in re-contacting store managers. Figure D.1 shows the resulting distribution of elapsed time across stores.

Moreover, our main empirical objective is to test Corollary 1 by studying the effects of large chain entry for small k (capacity-constrained) versus large k (less capacity-constrained) local stores. This focus on relative effects further mitigates concerns about market-specific time-varying shocks that affect all local stores, which would cancel out, in principle, when we compare the impact on capacity-constrained and less capacity-constrained local stores.

We estimate the following regressions at the store level:

$$\Delta Assortment_{i,t=2} = \alpha + \beta\, ChainEntry_{i,(t=1,t=2)} + X'_{i,t=1}\gamma + \varepsilon_i \quad (6)$$

where $\Delta Assortment_{i,t=2}$ measures changes in store i's assortment (e.g., product additions and removals explicitly for competition or demand reasons) between baseline ($t = 1$) and endline ($t = 2$). $ChainEntry_{i,(t=1,t=2)}$ is an indicator equal to 1 if a large chain store (supermarket or convenience chain) entered within a defined radius (1 km in our standard specification) of store i between baseline and endline; and 0 otherwise. Stores with no entry *within 2 km* serve as the control group for consistency across our various trade area definitions. $X_{i,t=1}$ is a vector of baseline store-level covariates, including store revenue, managerial practices, and formal registration status, used to improve precision. Our key parameter is β, reflecting the impact of chain entry on assortment adjustments.

We consider various measures of assortment change as dependent variables. First, we measure whether stores report adding new products specifically due to competition or customers as a binary dependent variable. Second, we measure whether stores report removing existing products, again explicitly for customer or competition reasons. As placebo tests, we separately examine product additions and removals reported for reasons unrelated to competition or customers (e.g., supplier issues, margin adjustments), expecting null results.

Third, we construct a specialization index, i.e., a measure of within-store directional specialization relative to the baseline core assortment, using our text data on what products were added and removed. Specifically, we use a Large Language Model (LLM) to assign each product reported by store managers into hierarchical product clusters corresponding to categories. Then, assortment changes can be compared against the primary categories stocked by the store at baseline (t=1) per their self-description and their top products by volume. We treat the union of these responses as the store's baseline primary category set. At endline (t=2), we do not re-enumerate the assortment. Instead, we observe lists of products added and removed since baseline. We classify each added or removed product into standard categories defined by the LLM-based classifier. Using these classifications, we compute a specialization measure as the log ratio of *assortment depth* – number of products divided by number of categories reported – at endline (t=2) versus baseline (t=1). This continuous measure decreases, for example, when the products removed by store managers belong to the same categories as those that were reported as the primary categories at baseline, as well as when products added belong to different categories than were not reported as the primary categories at baseline.

When computing the assortment depth, there are two intermediate steps: (i) we need to specify the level of category granularity and thus, to obtain a robust measure, we repeat the assortment depth calculation under three granularity levels (prompting the LLM to impose between 10 to 33 product categories in the data), and (ii) we fine-tune using a standard few-shot learning approach where we manually assign products to categories for a small subset of observations. We normalize the final index such that coefficients can be interpreted in terms of standard deviations. In addition to this index, we report results for the simpler assortment depth ratio measure.

6.2. Results

Table 4 reports estimates of equation 6 on binary measures of assortment changes made due to competition or customers (Columns 1 to 3) versus placebo reasons (Columns 4 to 6). We find large, positive, and statistically significant effects of chain entry on assortment changes for reasons related to competition or customers. For instance, the probability that such a product removal was made increases by 14 percentage points, a large relative increase over the likelihood that a control store made such a removal (base probability of 0.17). The probability that a product was added for reasons of competition or customers increases by 15 pp, again for a large relative effect of 45% (base probability of 0.33). In contrast, we observe null effects of large chain entry on assortment changes for placebo reasons, despite these placebo changes occurring with roughly equivalent probability in the data (0.42 vs. 0.41 base probability in the control group). This initial set of results builds our argument that store managers are responsive to large chain entry when it comes to assortment adjustments. We show that these results are robust to: (i)the definition of the trade area – they hold when we consider large chain entries within 2 km (Table D.1) and 0.5 km (Table D.2); (ii) the trimming of outliers – they hold without trimming (Table D.3) or trimming 5% of observations on right tail of time elapsed between t=1 and t=2 (Table D.4); and (iii) excluding non-food retailers from the analysis (Table D.5).

Table 4
Entry of Large Chains and Within-Firm Assortment Changes

	Assortment Changes: Competition/Customers			Assortment Changes: Placebo Reasons		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.14***	0.15***	0.19***	0.01	-0.07	-0.04
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Observations	388	388	388	388	388	388
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.17	0.33	0.42	0.22	0.26	0.41

Note: The table displays the assortment changes within stores resulting from the entry of large chains (within a 1 km radius), estimating equation 6 using a linear probability model with observations at the store level. Variation across columns denote diverse outcomes, respectively: assortment removal due to competition or customers, addition due to competition or customers, change due to competition or customers, removal due to other (placebo) reasons, addition due to other (placebo) reasons, and change due to other (placebo) reasons. The indicated regressions include store characteristic controls for precision (monthly profits, total employees, number of weekly customers, tax registration status, and an index of business practices). Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table 5 supports Corollary 1 in showing that chain entry drives assortment changes in stores that are more capacity-constrained. The median store in our data carried 60 to 80 SKUs at baseline, and thus we define a capacity-constrained store (small k) as one with up to 60 SKUs, in line with our model's conceptualization of k. Table 5 reports estimates of equation 6 for small k and large k stores. We find null effects, with point estimates close to zero, of chain entry on the likelihood of assortment changes for large k stores. The positive effects of chain entry on specialization are concentrated in the small k stores. Again, this pattern of results is not sensitive to: (i) the definition of trade area, holding firm when we consider large chain entries within 2 km (Table D.6) and 0.5 km (Table D.7); (ii) the trimming of outliers, holding without trimming (Table D.8) or trimming 5% of observations on right tail of time elapsed between t=1 and t=2 (Table D.9); and (iii) excluding non-food retailers from the analysis (Table D.10). Reassuringly, when we consider assortment changes for placebo reasons as the DV, small k and large k stores do not exhibit any differential pattern (Table D.11) in that across the board there are null effects of large chain entry.

Table 5
Entry of Large Chains and Within-Firm Assortment Changes by Store Capacity Constraints

	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.20***	0.24***	0.28***	-0.02	0.01	0.03
	(0.05)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)
Observations	239	239	239	149	149	149
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.10	0.33	0.40	0.34	0.31	0.47

Note: The table displays the assortment changes within stores resulting from the entry of large chains (within a 1 km radius), estimating equation 6 using a linear probability model with observations at the store level. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to competition or customers for more capacity-constrained stores, addition due to competition or customers for more capacity-constrained stores, change due to competition or customers for more capacity-constrained stores, removal due to competition or customers for less capacity-constrained stores, addition due to competition or customers for less capacity-constrained stores, and change due to competition or customers for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Finally, using our LLM-based specialization measures, we explicitly document that the stores facing nearby chain entry significantly increase their specialization relative to baseline. These results are presented in Table 6. The estimated effect of chain entry on our normalized specialization index is 0.21 SD. This effect is particularly pronounced among capacity-constrained local stores, precisely as Corollary 1 outlines. Column (2) shows the effect is statistically significant and positive for small k stores (0.28 SD), while Column (3) shows null effects with point estimates close to zero for large k stores. Columns (4)-(6) show results for the raw assortment depth ratio measure which divides observed products per category at t=2 by observed products per category at t=1. Again, we see the same pattern of results: the estimated effect of chain entry is 0.20 SD, and it is concentrated in small k stores (0.29 SD) with null effects for large k stores. As previously, these results are also robust to: (i) the definition of trade area, holding when we consider large chain entries within 2 km (Table D.12) and 0.5 km (Table D.13); (ii) the trimming of outliers, holding

Table 6
Entry of Large Chains and Within-Firm Specialization Changes by Store Capacity Constraints

	Specialization Index			Raw Depth Ratio (Std.)		
Dependant Variable:	Full	Small k	Large k	Full	Small k	Large k
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.21**	0.28**	0.08	0.20*	0.29**	0.02
	(0.10)	(0.12)	(0.22)	(0.10)	(0.12)	(0.22)
Observations	387	239	148	387	239	148
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	-0.14	-0.16	-0.07	-0.12	-0.16	-0.01

Note: The table displays the specialization within stores resulting from the entry of large chains (within a 1 km radius), estimated using OLS regressions with observations at the store level. The dependent variable is the specialization index described in Section 6.1. Variation across columns denote different subgroups, respectively: the full sample, more capacity-constrained stores, and less capacity-constrained stores. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

without trimming (Table D.14) or trimming 5% of observations on right tail of time elapsed between t=1 and t=2 (Table D.15); and (iii) excluding non-food retailers from the analysis (Table D.16).

In sum, our within-store findings sharply support our theoretical prediction that local stores strategically respond to increased competition from large chains by shifting assortments towards greater specialization, particularly among capacity-constrained stores.

7. Conclusion

How do small neighborhood stores survive in a retail world increasingly dominated by giants like Walmart and Amazon? Using nationwide microdata from the Mexican Statistics Institute (INEGI) and original primary surveys from hundreds of small offline retailers, we document a key survival strategy: defensive specialization.

Our findings speak to fundamental questions about the impact of e-commerce (and competition from large chains more generally) on local retail and consumer welfare. Waldfogel (2007)'s "tyranny of the majority" posits that substantial fixed costs lead local businesses to disproportionately serve mainstream consumers, with taste minorities benefiting primarily from large firms' scale advantages. Choi and Bell (2011) reinforce this view, showing that geographic variation in preference minority status explains variation in online sales penetration. We propose a contrasting equilibrium: as chains and large firms achieve dominance, local stores are forced to specialize in narrow niches. Thus, in equilibrium, taste minorities receive substantial service from small retailers – not despite, but because of large competitor dominance.

Our work also qualifies Anderson (2004)'s celebrated long tail theory, in that we argue that simply stocking bestsellers is becoming less viable for small local retailers post-entry of dominant competitors. Interestingly, Anderson's online long tail leverages *absence* of capacity constraints; our offline long tail arises precisely from their *presence*.

Our theory extends naturally to contemporary competitive dynamics beyond traditional retail. Platforms like Amazon, for example, are increasingly competing with third-party sellers through private labels such as Amazon Basics. How are third-party sellers adapting?

For small retailers, our findings suggest that specialization provides partial insurance against competition from larger alternatives. Pricing strategies should reflect the boutique effect – specialty stores can sustain premium margins that offset reduced traffic.

From a policy perspective, our welfare analysis reveals that the market equilibrium features *insufficient* specialization: when a local store is indifferent between remaining a generalist and becoming a specialist, consumers strictly prefer the latter. This systematic underspecialization arises because firms do not internalize the full consumer surplus gains from serving niche preferences more intensively. The implication is that policies encouraging small-store specialization – such as zoning regulations that protect specialty retailers or targeted support for niche-market pivots – may be welfare-improving, not merely redistributive. Market concentration may paradoxically increase retail diversity through induced specialization, but the equilibrium level of specialization still falls short of the social optimum.

Ultimately, our paper illuminates how small stores limit competitive damage. Defensive specialization allows partial adaptation to an increasingly concentrated retail landscape. As retail evolution continues, the question is not whether small stores can compete head-to-head with retail giants, but whether market structures permit sufficient specialization opportunities for diverse retail models to coexist.

References

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell (2019), "Food deserts and the causes of nutritional inequality," *The Quarterly Journal of Economics*, 134, 1793–1844.

Alter, Alexandra (2020), "Bookstores Are Struggling. Is a New e-Commerce Site the Answer?" *New York Times*, June 16.

Anderson, Chris (2004), "The Long Tail," *Wired Magazine*, October 1.

Anderson, Chris (2006), *The Long Tail: Why the Future of Business Is Selling Less of More*, Hachette UK.

ANTAD (2022), "Informe Anual 2022," Technical report, Asociacion Nacional de Tiendas de Autoservicio y Departamentales (ANTAD), Mexico, accedido el January 15, 2026.

Arcidiacono, Peter, Paul B Ellickson, Carl F Mela, and John D Singleton (2020), "The competitive effects of entry: Evidence from supercenter expansion," *American Economic Journal: Applied Economics*, 12, 175–206.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen (2017), "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan*," *The Quarterly Journal of Economics*, 132, 1101–1164.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro (2018), "Retail globalization and household welfare: Evidence from Mexico," *Journal of Political Economy*, 126, 1–73.

Bronnenberg, Bart J. and Paul B. Ellickson (2015), "Adolescence and the Path to Maturity in Global Retail," *Journal of Economic Perspectives*, 29, 113–134.

Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D Smith (2010), "The Longer Tail: The Changing Shape of Amazon's Sales Distribution Curve," SSRN 1679991.

Busso, Matías, O. Fentanes, and S. Levy (2018), "The Longitudinal Linkage of Mexico's Economic Census 1999-2014," *IDB Technical Note ; 1477*.

Caoui, El Hadi, Brett Hollenbeck, and Matthew Osborne (2024), "Dynamic Entry & Spatial Competition: An Application to Dollar Store Expansion," *Available at SSRN*.

Choi, Jeonghye and David R Bell (2011), "Preference Minorities and the Internet," *Journal of Marketing Research*, 48, 670–682.

Datta, Sumon and K Sudhir (2011), "The agglomeration-differentiation tradeoff in spatial location choice," *manuscript (Yale School of Management, Yale University, New Haven, Connecticut, USA)*.

David, H. A. and H. N. Nagaraja (2003), *Order Statistics*, Wiley, Hoboken, NJ, 3rd edition.

de Mel, Suresh, David McKenzie, and Christopher Woodruff (2008), "Returns to Capital in Microenterprises: Evidence from a Field Experiment*," *The Quarterly Journal of Economics*, 123, 1329–1372.

Economic Census (2019), “Censos Economicos 2019,” `https://www.inegi.org.mx/programas/ce/2019/`, accessed: 2025-07-02.

Ellickson, Paul B, Paul LE Grieco, and Oleksii Khvastunov (2020), “Measuring competition in spatial retail,” *The RAND Journal of Economics*, 51, 189–232.

ENIGH (2006), “Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2022,” Accessed: 2025-04-17.

Hoch, Stephen J, Eric T Bradlow, and Brian Wansink (1999), “The variety of an assortment,” *Marketing Science*, 18, 527–546.

Holmes, Thomas J. (2011), “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79, 253–302.

Hong, Sungtak, Kanishka Misra, and Naufel J. Vilcassim (2016), “The Perils of Category Management: The Effect of Product Assortment on Multicategory Purchase Incidence,” *Journal of Marketing*, 80, 34–52.

Hortaçsu, Ali and Chad Syverson (2015), “The Ongoing Evolution of US Retail: A Format Tug-of-War,” *Journal of Economic Perspectives*, 29, 89–112.

Hsieh, Chang-Tai and Benjamin A. Olken (2014), “The Missing "Missing Middle",” *Journal of Economic Perspectives*, 28, 89âĂŞ108.

Igami, Mitsuru (2011), “Does Big Drive out Small?” *Review of Industrial Organization*, 38, 1–21.

INEGI (2025), “Sistema de Clasificacion Industrial de America del Norte (SCIAN),” Accessed: 2025-07-02.

Jia, Panle (2008), “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, 76, 1263–1316.

Kravitz Hoeffner, Melissa (2022), “NYC’s Iconic Strand Bookstore Now Has a Coffee Shop,” *TimeOut New York*.

Lim, Stanley Frederick WT, Elliot Rabinovich, Sungho Park, and Minha Hwang (2021), “Shopping activity at warehouse club stores and its competitive and network density implications,” *Production and Operations Management*, 30, 28–46.

Mankiw, N. Gregory and Michael D. Whinston (1986), “Free Entry and Social Inefficiency,” *The RAND Journal of Economics*, 17, 48–58.

Martínez-de Albéniz, Victor, Diego Aparicio, and Jordi Balsach (2025), "The resilience of fashion retail stores," *Production and Operations Management*, 34, 1167–1187.

Maurice-Jones, Amelie (2025), "What's Behind New York's Bookstore Bar Boom?" *The Drinks Business*.

McKenzie, David (2020), "Small Business Training to Improve Management Practices in Developing Countries," *Development Research*.

McKenzie, David and Anna Luisa Paffhausen (2019), "Small Firm Death in Developing Countries," *The Review of Economics and Statistics*, 101, 645–657.

McKenzie, David and Christopher Woodruff (2017), "Business Practices in Small Firms in Developing Countries," *Management Science*, 63, 2967–2981.

Raffaelli, Ryan (2020), "Reinventing retail: the novel resurgence of independent bookstores," *Harvard Business School*, 20–068.

Talamas Marcos, Miguel Angel (2024), "Surviving Competition: Neighbourhood Shops versus Convenience Chains," *The Review of Economic Studies*, 92, 553–585.

Waldfogel, Joel (2007), *The Tyranny of the Market: Why You Can't Always Get What You Want*, Harvard University Press, Cambridge, Mass.

APPENDIX

A. Theoretical model extensions

As mentioned earlier, our basic model is based on a series of simplifying assumptions: we consider competition between *one* small store and *one* large entrant; we assume away price competition to zoom in on stocking decision; we assume individual consumers have preferences for *one* of two (symmetric) product categories; and we assume an *exogenously given* preference for small stores.

In this Appendix, we show that our main results, and their intuition, do not rely on these simplifying assumptions. In particular, in our next four Propositions, we consider what happens with asymmetric product categories; endogenous prices; eclectic consumers, who value both categories; competition between two symmetric stores (as well as with a larger chain entrant); and endogenous investments in amenities.

The goal of this exercise is twofold. First, we show that our *defensive specialization* results are not an artifact of the simplified framework we use to derive Propositions 1 and 2. Second, our model extensions allow us to derive a series of interesting additional results regarding the complementarity of specialization decisions and other strategic choices available to local stores.

A.1. Niche Product Categories

So far, we have assumed that both product category x and product category y have the same popular appeal. A more realistic case has one of the product categories – say, product category x – be a popular product category, whereas y is a less popular one – a niche product category. Suppose that there is a measure 1 of potential buyers, α of which are only interested in product category x; and suppose that $\alpha > \frac{1}{2}$. (So far, we have implicitly assumed that $\alpha = \frac{1}{2}$.) Consistent with the assumption that product categories x and y have different popular appeal, we assume that a fraction $\alpha\, z$ of the total products are of product category x, and a fraction $(1 - \alpha)\, z$ are of product category y.

Proposition 1 states that, as z increases, store b optimally switches from a general to specialty store. The next proposition complements that result by stating that, within the specialty strategy, store b optimally chooses the niche product strategy if z is high enough.

Proposition 3. *There exists a z_{xy} such that an active store b specializes in a niche product category (rather than a popular product category) if $z > z_{xy}$.*

Similar to Proposition 1, Proposition 3 has a dual interpretation where the comparative statics can be done with respect to k rather than z:

Corollary 3. *There exists a k_{xy} such that an active store b specializes in a popular product category if $k > k_{xy}$.*

Figure 5
Store profits from specializing in popular product category (π_x) or niche product category (π_y) as a function of z when $F(\widetilde{v}) = \widetilde{v}/v$.

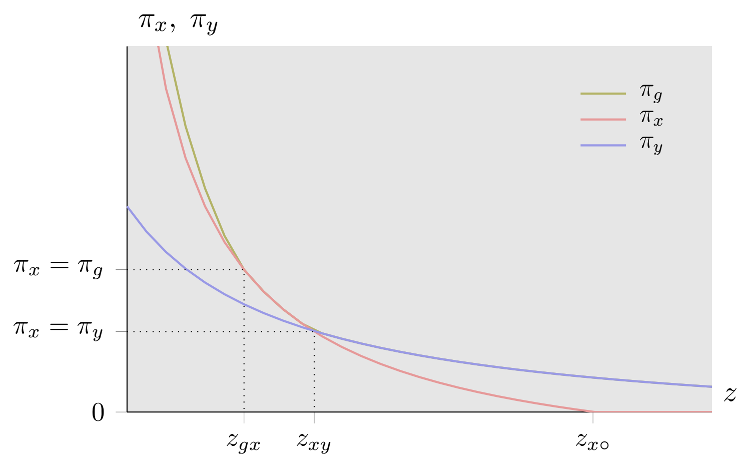


Figure 5 illustrates Proposition 3 and Corollary 3. The key insight is that, *relatively* speaking, a niche-product category store suffers less from an increase in z than a popular-product category store, in a way that is similar to, but different from, the general-specialty trade-off considered in Proposition 1. For low values of z, the advantage of a niche-product category store, in terms of higher intensive margin, is outweighed by the simple fact that a popular product category is more popular, that is, attracts a greater number of potential customers. For high values of z, however, the niche strategy becomes increasingly attractive, as illustrated by Figure 5. Specifically, for $z > z_{xy}$, π_y, the profit from a niche-product category strategy, is greater than π_x, the profit from a popular-product category strategy.

Formally, the proof of Proposition 3 proceeds by deriving the value z_x when $\pi_x = 0$ and establishing that, at that value, $\pi_y > 0$. This proof strategy is similar to that of Proposition 1. There is one difference, however: In Proposition 1, we show that $z > z_{gs}$ is a necessary and sufficient condition for specialization. By contrast, in Proposition 3 $z > z_{xy}$ is only a sufficient condition. The difference stems from the fact that we can prove the monotonicity of $\pi_s - \pi_g$ in general terms but not the monotonicity of $\pi_y - \pi_x$. If we further assume that v is uniformly distributed, then the condition $z > z_{xy}$ becomes a necessary and sufficient condition.[15]

A.2. Endogenous Prices

So far, we have assumed that all items are priced \$1. We now explicitly consider pricing choices. Our goal is to verify the robustness of our previous findings as well as to develop additional intuition regarding the comparative statics of the chain's expansion.

Recall that the actual market structure we have in mind includes one dominant firm and a large number of fringe firms. Although for simplicity we focus on the decisions of one representative fringe firm, it makes sense to treat firms a and b as different types of strategic players. Consistent with this interpretation, we assume that firm a acts a price leader by setting p_a first.

Given p_a, the local store b responds by deciding whether to be a general store or a specialty store as well as by setting its price, which we denote by p_g if the store is a general store and p_s if the store is a specialty store. Our focus is on firm b's decisions. Accordingly, we take p_a as an exogenous variable (and later consider comparative statics with respect to it).[16] Similar to Propositions 1 and 3, we make a parametric assumption so as to eliminate trivial corner solutions (if the assumption below fails to hold, then choosing to be a specialty store is always optimal).

Our next result extends the main intuition of Proposition 1, adding one new dimension of comparative statics.

Proposition 4. *Suppose that $p_a > \underline{p}$, where*

$$\underline{p} = w + \frac{m(k) - \sqrt{2}\, m(k/2)}{\sqrt{2} - 1}$$

Then, there exists a threshold z_{gs} such that store b optimally chooses to be a specialty store if $z > z_{gs}$. In the right neighborhood of z_{gs}, the specialty store sets a higher price, captures a lower market share and earns a higher profit than a general store.

When discussing Proposition 1, we argued that the trade-off between a general and a specialty store is a trade-off between the extensive margin (which favors a general store) and the intensive margin (which favors a specialty store). The proof of Proposition 4 establishes that, when it comes to price setting, only the intensive margin matters. This explains why a specialty store sets a higher price than a general store. By devoting its space to one product

15. The proof can be obtained from the authors upon request.
16. Endogenizing the chain's decisions would be a promising direction for future research.

category only, a specialty store elicits a higher willingness to pay from buyers interested in that product category, which in turn allows the store to set higher prices. This in turn increases the store's incentives to specialize.

Similar to Proposition 1, Proposition 4 establishes that, if chain a is big enough (high z), then firm b is better off by becoming a specialty store. The main intuition for the z-threshold part of Proposition 4 is similar to Proposition 1: As total supply z increases, the specialty store option becomes *relatively* more attractive. In sum, the first part of Proposition 4 shows that the intuition from Proposition 1 is robust to the introduction of pricing.

The novel aspect of Proposition 4 is its second part, the statement that, past the disruption level z_{gs}, a specialty store sets a higher price, captures a *lower* market share and earns a higher profit than a general store. We call this the *boutique effect*. The specialty store in the model with fixed prices trades-off extensive margin and intensive margin so as to maximize the number of customers. By switching from general to specialty store, firm b loses potential customers, but its offering becomes so much more attractive to its reduced set of customers that it ends up attracting more customers. By contrast, once we introduce prices we observe that the switch to a specialty-store strategy not only sacrifices *potential* demand but also sacrifices *actual* demand. Such drop in actual demand is more than compensated by an increase in the intensive margin via higher retail prices.

A.3. Eclectic Consumers

So far, we have assumed that consumers are divided into x fans and y fans. Specifically, the value v of an item outside of a consumer's preferred product category is zero. At the opposite extreme, consider the case when consumers are totally eclectic, that is, they value both product categories equally.

Clearly, eclectic consumers are bad news for specialty stores. Before, an x fan valued a specialty store at $m(k)$ and the chain store at $m(z/2)$. By contrast, an eclectic consumer values the chain store at $m(z)$ whereas the specialty store is still valued at $m(k)$ (here we are excluding the preference parameter $\widetilde{w}$).

Regarding a general store, the analysis is not as obvious. Before, the value of a general store was $m(k/2)$ for an x fan or a y fan, whereas the value of the chain was $m(z/2)$. By contrast, an eclectic consumer values the chain at $m(z)$ whereas the general store is valued at $m(k)$ (again, we are excluding the preference parameter $\widetilde{w}$). In which case is the general store better off? The answer depends on which difference is greater, $m(z/2) - m(k/2)$ or $m(z) - m(k)$. Notice that $m(z) - m(k) > m(z/2) - m(k/2)$ if and only if $m(z) - m(z/2) > m(k) - m(k/2)$. Since $z > k$, we have $z - z/2 > k - k/2$, which in turn suggests the inequality holds. However, concavity of $m(t)$ would work against the inequality. Suppose that $F = v$ is linear, so that $m(t) = t/(1+t)$. Then the function $m(t) - m(t/2)$ is non-monotonic, first increasing for $t \in [0, \sqrt{2}]$ and then decreasing. This implies that we can find values of z and k such that the inequality is, in turn, true or false. So, even assuming a specific distribution of v, we cannot guarantee that a general store is better off or worse off when serving eclectic consumers rather than polarized consumers.

It has long been argued that large chains benefit from increased consumer specialization due to their larger size, since local stores cannot, due to their limited size, cater to each consumer's idiosyncrasies (Anderson (2004), Waldfogel (2007)). However, as the above analysis shows, this is not necessarily true when we endogenize local stores' strategies: more specialized consumers allow specialty stores to emerge, which can be detrimental to the chain's profits.

A.4. Traditional Store Amenities

We know that, *ceteris paribus*, local stores' survival is crucially dependent on the relative consumer preferences for traditional retail ($\widetilde{w}$). So far, we have treated this distribution as exogenous. We now consider the case when stores make investments that determine consumer preferences. For example, local shops might build relationships with the community. The owner, often a neighbor, can offer tailored customer service and engage in conversations. The store itself may be a gathering point for the neighbors. Moreover, knowing the neighbors so well allows local shops to offer amenities unfeasible for chains, such as informal credit. These features are appealing because chains cannot directly replicate them.

Are amenities a complement or a substitute for specialization? And which stores benefit the most from them? The following result provides an answer:

Proposition 5. *Generalist stores have a higher incentive to invest in amenities:* $\partial \pi_g / \partial w > \partial \pi_s / \partial w > 0$. *This incentive gets stronger the larger the size of the entrant:* ($\partial^2 \pi / \partial w \, \partial z > 0$) *and/or the smaller the size of the store* ($\partial^2 \pi / \partial w \, \partial k < 0$).

The first part of Proposition 5 states that generalist stores get more bang for the buck when it comes to investment in local amenities. The intuition is that, by serving a wider customer base, the investment can be spread on a larger number of consumers.

The second part of Proposition 5 refers to cross-partial derivatives, that is, to moderating factors in evaluating the benefits from investment in amenities. First, the greater the chain is, the more a local store benefits from investment. Intuitively, just as specialization is a strategic reaction to a large chain entry, so is investment in store amenities. Second, the greater the local store, the lower the store's benefit from investment in amenities. Similar to Propositions 1 and 3, these two results (comparative statics with respect to z and to k are connected: ultimately, investment in local store amenities is more profitable the greater the gap between local store size (k) and chain size (z).

Proposition 5 suggests that, for a small store (low k), investing in amenities (i.e., increasing the value of w) may provide an alternative strategy to specialization. This is particularly the case when a significant fraction of consumers are eclectic (so that specialization is not a profitable strategy).

A.5. Local Store Competition

Up to now, we have considered competition between one large chain and one local store. Implicitly, the idea is that there are a plethora of small (possibly independent) local stores with a catchment area that does not overlap with any other local store. Consider now the case when two local stores, say b_0 and b_1, do compete for the same potential demand. Specifically, we assume a consumer is characterized by a local-store preference $\widetilde{w}$ and a relative preference between stores b_0 and b_1 in the form of a location $\widetilde{d} \in [0, 1]$ and transportation cost τ per unit of distance to store b_0 (located at 0) and to store b_1 (located at 1). Moreover, we assume that $\widetilde{d}$ and $\widetilde{w}$ are independently and uniformly distributed: $\widetilde{d} \sim U[0, 1]$ and $\widetilde{w} \sim U[0, w]$. Our main result is that, under competition, the choices of product category by stores b_0 and b_1 exhibit strategic complementarities.

Proposition 6. *Let z be such that store b_0 and b_1 are indifferent between a general- and a specialty-store strategy absent competition between local stores. In the neighborhood of z, being a specialty store is a strict best response to the rival choosing to be a specialty store.*

Proposition 6 suggests that competition provides an additional force pushing in the direction of specialization. Suppose that we fix firm b_1's strategy at being a general store. As z crosses a certain threshold, say $z_\circ$, then firm b_0's optimal strategy switches to becoming a specialty firm (of either x or y). However, if firm b_1 has become a specialty firm (choosing, say, product category y), then, *even if z is lower than $z_\circ$* (by a little), firm b_1 also optimally switches to being a specialty (specializing in the niche that firm b_1 did not).

A.6. Welfare Analysis

All of our analysis so far has focused on firm b's profits and optimal choices. A natural follow-up question is the relation between firm b's decisions and consumer welfare. Let us go back to the model with fixed prices and one local store, firm b. Let us consider, as in the initial model, the choice between being a general and being a specialty store. Suppose social welfare is given by consumer surplus plus firm profits. Since all sellers set $p = 1$ and the market is covered (all consumers make a purchase), consumer surplus is a sufficient statistic of social welfare.

Figure 6

Firm profit and consumer welfare. Effects of switching from general to specialty x store.

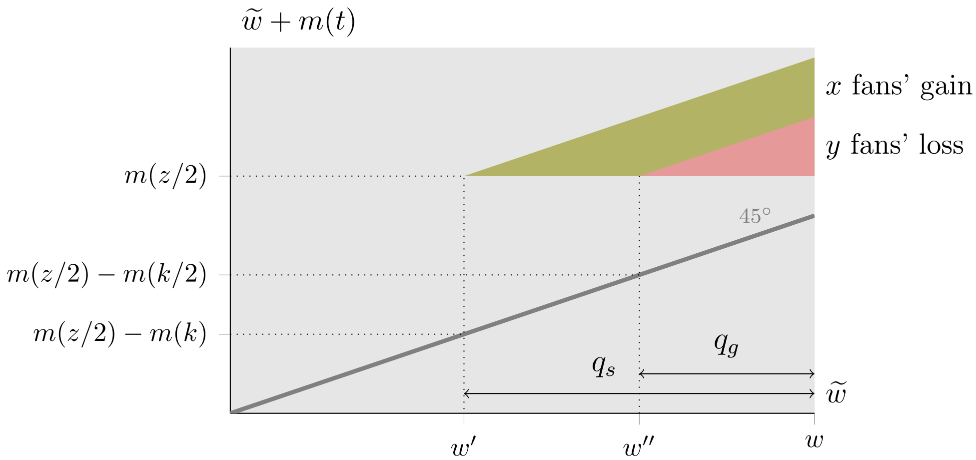


Figure 6 illustrates the contrast between a general and a specialty store when competing against firm a. On the horizontal axis we measure each consumer's value of $\widetilde{w}$, that is, their disutility from buying from firm a. On the vertical axis we measure the advantage, in terms of vertical quality, of the chain store with respect to the local store. The 45° line measures the points at which the "horizontal" differentiation advantage of firm b exactly compensates the "vertical" differentiation advantage of firm a.

Consider first the case of a general store b. Its disadvantage with respect to store a is given by $m(z/2) - m(k/2)$. It follows that only consumers with a value of $\widetilde{w}$ greater than w'' purchase at the local store. Since $\widetilde{w}$ is uniformly distributed, we conclude that firm b's market share is given by $q_g = (w - w'')/w$.

Consider now the case of a specialty store b. Its disadvantage with respect to store a is given by $m(z/2) - m(k)$. It follows that only consumers with a value of $\widetilde{w}$ greater than w' purchase at the local store. Since $\widetilde{w}$ is uniformly distributed, we conclude that firm b's market share (among its product category followers) is given by $q_s = (w - w')/w$. However, we must keep in mind that if firm b focuses on product category x, for example, then it loses potential buyers who are only interested in y. In other words, by becoming a specialty store firm b halves its potential demand. Therefore, its market share is $(w - w')/(2w)$.

The values of z and k were selected so that $\pi_g = (w - w'')/w = (w - w')/(2w) = \pi_s$. In other words, for the particular values of z and k underlying Figure 6, firm b is indifferent between being a general store or being a specialty store. Consumers, however, are not indifferent between the two types of store. Consumer surplus is given by the area below

$$\max\{m(z/2), \widetilde{w} + m(\widetilde{k})\}$$

where $\widetilde{k} = k/2$ or $\widetilde{k} = k$ for a general and a specialty store, respectively. It follows that, for product category x consumers, the switch from a general to a product category x specialty store implies an increase in consumer surplus given by the green trapezoid in Figure 6. By contrast, for product category y consumers the switch implies a decrease in consumer surplus given by the red area in Figure 6. By construction, the green area is greater than the red area. More generally, we have just established the following result:

Proposition 7. *When store b is indifferent between being a general or a specialty store, the average consumer strictly prefers the latter.*

Intuitively, consumer surplus is “convex” in the vertical utility provided by the local store. This implies that consumers prefer the “bet” of having a specialty store of their preferred product category with probability 50% than a general store with probability 100%.

This intuition is related to a number of results in the IO literature. Mankiw and Whinston (1986) provide conditions such that, in equilibrium, there is excess entry into a market. Intuitively, the entrant does not correctly take into account the positive externality it creates for consumers nor the negative externality it creates for its competitors. Similarly, our firm b does not take into account the positive surplus effect it has on the consumers who like the product category in which they specialize.

ONLINE APPENDIX

A. Proofs of Theoretical Results

Proof of Proposition 1: Consider the case of a general store. For a x (or y) consumer, visiting b yields expected value

$$\widetilde{w} + m(k/2)$$

By contrast, buying at a yields expected value

$$m(z/2)$$

given that half of the total products correspond to product category x (or y). The indifferent buyer is characterized by

$$\widetilde{w} = m(z/2) - m(k/2)$$

whenever $m(z/2) - m(k/2) < w$. (Otherwise, every consumer strictly prefers seller a and b makes zero profits.) Finally, b's expected profit (when strictly positive) is given by

$$\pi_g = 1 - \big(m(z/2) - m(k/2)\big)\,/w \tag{7}$$

Consider now the case of a store specializing in product category x. For an x consumer, visiting b yields expected value

$$\widetilde{w} + m(k)$$

For a y consumer, the value of the x specialty store is zero. As before, buying at a yields expected value

$$m(z/2)$$

both for x and for y consumers. The indifferent x buyer is now characterized by

$$\widetilde{w} = m(z/2) - m(k)$$

whenever $m(z/2) - m(k) < w$. (Otherwise, every consumer strictly prefers seller a and b makes zero profits.) Finally, b's expected profit (when strictly positive) is given by

$$\pi_s = \tfrac{1}{2}\left(1 - \big(m(z/2) - m(k)\big)\,/w\right) \tag{8}$$

(Note that, by specializing, b expects to make, at most, $\frac{1}{2}$ in sales. This is because it will have lost all potential consumers from the product category it did not specialize in.)

As $z \to 2k^+$ (the smallest permissible chain size given $k < z/2$), being a general store yields

$$\pi_g \to \big(1 - \big(m(k) - m(k/2)\big)/w\big)$$

whereas being a specialist ensures profits

$$\pi_s \to \tfrac{1}{2}\left(1 - \big(m(k) - m(k)\big)/w\right) = \tfrac{1}{2}.$$

Intuitively, when $z \to 2k^+$, the specialist store has the same depth in its genre of choice as the chain. The condition ensures that it captures all of its potential demand, that is, 1/2.

We have that $\pi_g > \pi_s$ whenever $(m(k) - m(k/2))/w < 1/2$, or equivalently when $w > 2(m(k) - m(k/2))$. This holds by the lower bound in Assumption 3. Therefore, being a generalist dominates being a specialist when $z \to 2k^+$.

At the opposite end, let z_g be such that $\left(m(z_g/2) - m(k/2)\right)/w = 1$. For $z = z_g$, we have $\pi_g = 0$, whereas

$$\pi_s = \tfrac{1}{2}\left(1 - \big(m(z_g/2) - m(k)\big)/w\right) > \tfrac{1}{2}\left(1 - \big(m(z_g/2) - m(k/2)\big)/w\right) = 0$$

Such a z_g exists and is finite whenever $\lim_{z\to\infty}\big(m(z/2) - m(k/2)\big)/w > 1$, which is implied by $w < v - m(k/2)$, the upper bound in Assumption 3.

Given continuity of π_g and π_s, it follows from the intermediate value theorem that there exists an $z_{gs} \in (2k, z_g)$ such that $\pi_g(z_{gs}) = \pi_s(z_{gs})$, where for notational simplicity we have suppressed the store profit's dependence on k and w. To show that z_{gs} is unique we note that

$$\frac{d(\pi_s - \pi_g)}{dz} = \big(-m'(z/2) + 2\,m'(z/2)\big)/(4\,w) = m'(z/2)/(4\,w) > 0 \tag{9}$$

where the inequality follows from the fact that $m(z)$ is strictly increasing for every z. This concludes the first part of the proof.

To show that $z_{gs}(k, w)$ increases in k and w, we compute the derivative of the profit difference $(\pi_s - \pi_g)$ with respect to k and w:

$$\frac{\partial(\pi_s - \pi_g)}{\partial k} = \frac{m'(k)}{2\,w} - \frac{m'(k/2)}{2\,w} = \frac{1}{2\,w}\big(m'(k) - m'(k/2)\big) < 0 \tag{10}$$

where the inequality follows from concavity of m. Similarly,

$$\frac{\partial(\pi_s - \pi_g)}{\partial w} = \frac{m(z/2) - m(k)}{2\,w^2} - \frac{m(z/2) - m(k/2)}{w^2} = (\tfrac{1}{2} - \pi_s)/w - (1 - \pi_g)/w$$

where the second equality follows from (7) and (8). By definition, $\pi_s = \pi_g = \overline{\pi}$ at $z = z_{gs}$.

It follows that

$$\left.\frac{\partial(\pi_s-\pi_g)}{\partial w}\right|_{z\,=\,z_{gs}} = (\tfrac{1}{2}-\overline{\pi})/w-(1-\overline{\pi})/w=-1/(2\,w)<0 \tag{11}$$

By the implicit function theorem,

$$\frac{\partial z_{gs}(k,w)}{\partial k}=-\frac{\partial(\pi_s-\pi_g)/\partial k}{\partial(\pi_s-\pi_g)/\partial z}>0$$

where the inequality follows from (9) and (10). Also by the implicit function theorem,

$$\left.\frac{\partial z_{gs}(k,w)}{\partial w}\right|_{z\,=\,z_{gs}}=-\frac{\partial(\pi_s-\pi_g)/\partial w\,|_{s\,=\,z_{gs}}}{\partial(\pi_s-\pi_g)/\partial z}>0$$

where the inequality follows from (9) and (11).

■

Proof of Corollary 1:

We have that

$$\frac{\partial(\pi_g-\pi_s)}{\partial k}=\frac{1}{2}m'(k/2)-\frac{1}{2}m'(k)>0$$

by concavity of m. Moreover, as $k\to z/2^-$ (the largest permissible local store size given $k<z/2$), we have $\pi_g\to 1-(m(z/2)-m(z/4))/w$ and $\pi_s\to 1/2$. By the same argument as in the proof of Proposition 1, when $k\to z/2^-$ (equivalently, $z\to 2k^+$), the lower bound in Assumption 3 ensures $\pi_g>\pi_s$.

Conversely, we know that $\pi_g=0$ whenever $m(z/2)-m(k/2)\geq w$, while $\pi_s=0$ whenever $m(z/2)-m(k)\geq w$. Denote by k_g^* and k_s^* the two values of k that satisfy these two with equality. Because both expressions are decreasing in k, these exist and are non-negative if and only if $m(z/2)\geq w$.

Now notice that, by concavity of m, we have $k_g^*=2k_s^*$. Thus, whenever k_g^* and k_s^* are positive, we have that $k_g^*>k_s^*$. Since profits are truncated at zero when the indifferent consumer lies above w, we have

$$\pi_s\geq\pi_g=0,\quad \forall k\in[k_s^*,\ k_g^*],$$

with strict inequality for $k\in(k_s^*,\ k_g^*]$.

Combining our observations, we have that the difference $\pi_g-\pi_s$ is negative for $k\in[k_s^*,\ k_g^*]$ and monotonically increases, becoming strictly positive for $k\to z/2^-$. Thus, there exists a unique k_{gs} such that $\pi_s(k_{gs}, z) = \pi_g(k_{gs}, z)$.

Now, we want to show that $k_{gs}(z, w)$ is increasing in z and decreasing in w. To do so, we appeal to the Implicit Function Theorem again, in a similar fashion as in the proof of Proposition 1.

We have that

$$\frac{\partial k_{gs}(z,w)}{\partial z} = -\frac{\partial(\pi_g - \pi_s)/\partial z}{\partial(\pi_g - \pi_s)/\partial k} > 0$$

and

$$\frac{\partial k_{gs}(z,w)}{\partial w} = -\frac{\partial(\pi_g - \pi_s)/\partial w}{\partial(\pi_g - \pi_s)/\partial k} < 0,$$

which concludes the proof.■

Proof of Proposition 2: Define the variable profits (before fixed costs) as

$$\overline{\pi}_g = 1 - \big(m(z/2) - m(k/2)\big)/w, \quad \overline{\pi}_s = \tfrac{1}{2}\Big(1 - \big(m(z/2) - m(k)\big)/w\Big)$$

so that profits net of fixed costs are $\pi_g = \overline{\pi}_g - ck$ and $\pi_s = \overline{\pi}_s - ck$.

A general store exits when $\pi_g < 0$, i.e., when $c > \overline{\pi}_g/k$. Similarly, a specialty store exits when $c > \overline{\pi}_s/k$. Since c is drawn from a distribution $G(c)$ independent of store type, the probability that a general store exits is $1 - G(\overline{\pi}_g/k)$ and the probability that a specialty store exits is $1 - G(\overline{\pi}_s/k)$.

General stores are more likely to exit than specialty stores if and only if $\overline{\pi}_g < \overline{\pi}_s$, or equivalently $\overline{\pi}_s - \overline{\pi}_g > 0$.

From the proof of Proposition 1, we know that $\overline{\pi}_s - \overline{\pi}_g$ is strictly increasing in z, and there exists z_{gs} such that $\overline{\pi}_s = \overline{\pi}_g$. For $z > z_{gs}$, we have $\overline{\pi}_s > \overline{\pi}_g$, which implies that specialty stores have higher variable profits and therefore a lower exit threshold. Since G is a CDF (weakly increasing), $\overline{\pi}_s > \overline{\pi}_g$ implies $G(\overline{\pi}_s/k) > G(\overline{\pi}_g/k)$, so specialty stores have a strictly lower probability of exit.

Therefore, if $z > z_{gs}$ (i.e., "z is high enough"), general stores are more likely to exit than specialty stores. ■

Proof of Proposition 3: Suppose store b specializes in product category x, the popular product category ($\alpha > \frac{1}{2}$). Then store b reaches at most α of its potential customers. The indifferent customer (indifferent between store a and store b) has $\widetilde{w}$ such that

$$m(\alpha\, z) = m(k) + \widetilde{w}$$

where $\alpha\, z$ is total supply of products of product category x, all of which are available at store a; and k is the supply of products of product category x at store b (in other words, all of store b's capacity, k, is devoted to carrying product category x products). It follows that, of all store b potential customers, a fraction α is interested in the product category offered by store b, and a fraction $\big(m(\alpha\, z) - m(k)\big)\,/w$ of this fraction prefers store a to store b. This implies that store b's profit from specializing in product category x is given by

$$\pi_x = \alpha\Big(1 - \big(m(\alpha\, z) - m(k)\big)\,/w\Big)$$

Similarly, the profit from specializing in product category y is given by

$$\pi_y = (1-\alpha)\bigg(1 - \Big(m\big((1-\alpha)\, z\big) - m(k)\Big)\,/w\bigg)$$

If $z = 0$, that is, if the chain is out of the picture, then the popular product category x is trivially a dominant strategy: the store sells to a measure α of consumers, whereas the niche-product category store sells to a measure $1 - \alpha < \alpha$ only (and at the same price). At the opposite end, let z_x be the value of z such that $\pi_x = 0$. Such an z will exist whenever $\lim_{z\to\infty}\big(m(\alpha\, z) - m(k)\big)\,/w > 1$, which is equivalent to the condition in the Proposition. We then have

$$\pi_y = (1-\alpha)\bigg(1 - \Big(m\big((1-\alpha)\, z_x\big) - m(k)\Big)\,/w\bigg) > \alpha\Big(1 - \big(m(\alpha\, z_x) - m(k)\big)\,/w\Big) = 0$$

(If this condition does not hold — for instance because w or k are very large, or $m(n)$ is very flat —, then it may always be optimal for the store to choose the popular product category.)

Given continuity of π_x and π_y, the intermediate value theorem implies that there exists at least one value $\widehat{z}_{xy} \in (0, z_x)$ such that $\pi_x(\widehat{z}_{xy}) = \pi_y(\widehat{z}_{xy})$, where for notational simplicity we have suppressed the store profit's dependence on k and w. Let z_{xy} be the highest of these values. Then $\pi_y \geq \pi_x$ for $z > z_{xy}$.

Consider now the comparative statics with respect to k. Both π_x and π_y are increasing in k (since $m'(k) > 0$). As k becomes large relative to $(1-\alpha)z$, the specialty store in category y eventually stocks all available y products, so further increases in k only benefit the x-specialist. More precisely, for $k > (1-\alpha)z$, we have $\pi_y = (1-\alpha)[1 - (m((1-\alpha)z) - m((1-\alpha)z))/w] = 1-\alpha$, which is constant in k, while π_x continues to increase. It follows that there exists k_{xy} such that $\pi_x > \pi_y$ for $k > k_{xy}$. ■

Proof of Proposition 4: We first solve for the optimal prices of a general store given that store a sets p_a. Store g's profit is given by $\pi_g = p_g\, q_g$, where q_g, the store's sales, are given by

$$q_g = 1 - \big(m(z/2) - m(k/2) - p_a + p_g\big)\,/w$$

The profit-maximizing price, quantity and profit levels are given by

$$\widehat{p}_g = \tfrac{1}{2}\big(w - m(z/2) + m(k/2) + p_a\big) \quad (12)$$
$$\widehat{q}_g = \tfrac{1}{2}\big(w - m(z/2) + m(k/2) + p_a\big)\,/w = \widehat{p}_g/w \quad (13)$$
$$\widehat{\pi}_g = \widehat{p}_g\,\widehat{q}_g = (\widehat{p}_g)^2/w \quad (14)$$

In the case of a specialty store, profit is given by $\pi_s = p_s\, q_s$, where q_s, the store's sales, are given by

$$q_s = \tfrac{1}{2}\Big(1 - \big(m(z/2) - m(k) - p_a + p_s\big)\,/w\Big)$$

The profit-maximizing price, quantity and profit levels are given by

$$\widehat{p}_s = \tfrac{1}{2}\big(w - m(z/2) + m(k) + p_a\big) \quad (15)$$
$$\widehat{q}_s = \tfrac{1}{4}\big(w - m(z/2) + m(k) + p_a\big)\,/w = \widehat{p}_s/(2\,w) \quad (16)$$
$$\widehat{\pi}_s = \widehat{p}_s\,\widehat{q}_s = (\widehat{p}_s)^2/(2\,w) \quad (17)$$

Direct inspection of (12) and (15) reveals that

$$\widehat{p}_s > \widehat{p}_g$$

that is, in equilibrium specialty stores set a higher price. Moreover, from (12)–(13) and (15)–(16) we conclude that

$$\widehat{p}_s/\widehat{q}_s = 2\,w > \widehat{p}_g/\widehat{q}_g = w \quad (18)$$

Consider the extreme case when $z = 0$. Straightforward computation shows that $\widehat{\pi}_g > \widehat{\pi}_s$ if the condition in the Proposition holds. At the opposite end, let z_g be such that $\widehat{p}_g = 0$. Comparing (12) and (15), we see that, at $z = z_g$, $\widehat{p}_s > \widehat{p}_g = 0$. From (14) and (17) we conclude that, at $z = z_g$, $\widehat{\pi}_s > \widehat{\pi}_g = 0$. Since both $\widehat{\pi}_s$ and $\widehat{\pi}_g$ are continuous we conclude by the intermediate-value theorem that there exists at least one $\widetilde{z}_{gs}$ such that $\widehat{\pi}_s = \widehat{\pi}_g$. Let z_{gs} be the highest of these values. Then $\widehat{\pi}_s > \widehat{\pi}_g$ when $z_{gs} < z < z_g$.

Finally, notice that, at $z = z_{gs}$, $\widehat{\pi}_g = \widehat{\pi}_s$, that is, $\widehat{p}_g\,\widehat{q}_g = \widehat{p}_s\,\widehat{q}_s$. Since, from (18),

$\widehat{p}_s/\widehat{q}_s > \widehat{p}_g/\widehat{q}_g$, it must be that, at $z = z_{gs}$, $\widehat{p}_s > \widehat{p}_g$ and $\widehat{q}_s < \widehat{q}_g$. Since these are strict inequalities, they also hold in the neighborhood of $z = z_{gs}$. It follows that, in the right neighborhood of $z = z_{gs}$, a specialty store earns a higher profit, sets a higher price, and captures a lower market share. ■

Proof of Proposition 5: The proof follows straightforwardly from the definitions of π_g and π_s. Specifically, from (7) and (8) we derive

$$\begin{aligned} \frac{\partial \pi_g}{\partial w} &= \frac{m(z/2) - m(k/2)}{w^2} \\ \frac{\partial \pi_s}{\partial w} &= \frac{m(z/2) - m(k)}{2\,w^2} \end{aligned} \tag{19}$$

Since $m(k) > m(k/2)$ by strict monotonicity of m, we have $m(z/2) - m(k) < m(z/2) - m(k/2)$, which implies

$$0 < \frac{\partial \pi_s}{\partial w} < \frac{\partial \pi_g}{\partial w}$$

Taking derivatives of (19) with respect to z, we get

$$0 < \frac{\partial^2 \pi_s}{\partial w\ \partial z} < \frac{\partial^2 \pi_g}{\partial w\ \partial z}$$

Taking derivatives of (19) with respect to k, we get

$$\frac{\partial^2 \pi_g}{\partial w\ \partial k} = \frac{-m'(k/2)/2}{w^2} < 0, \qquad \frac{\partial^2 \pi_s}{\partial w\ \partial k} = \frac{-m'(k)}{2\,w^2} < 0$$

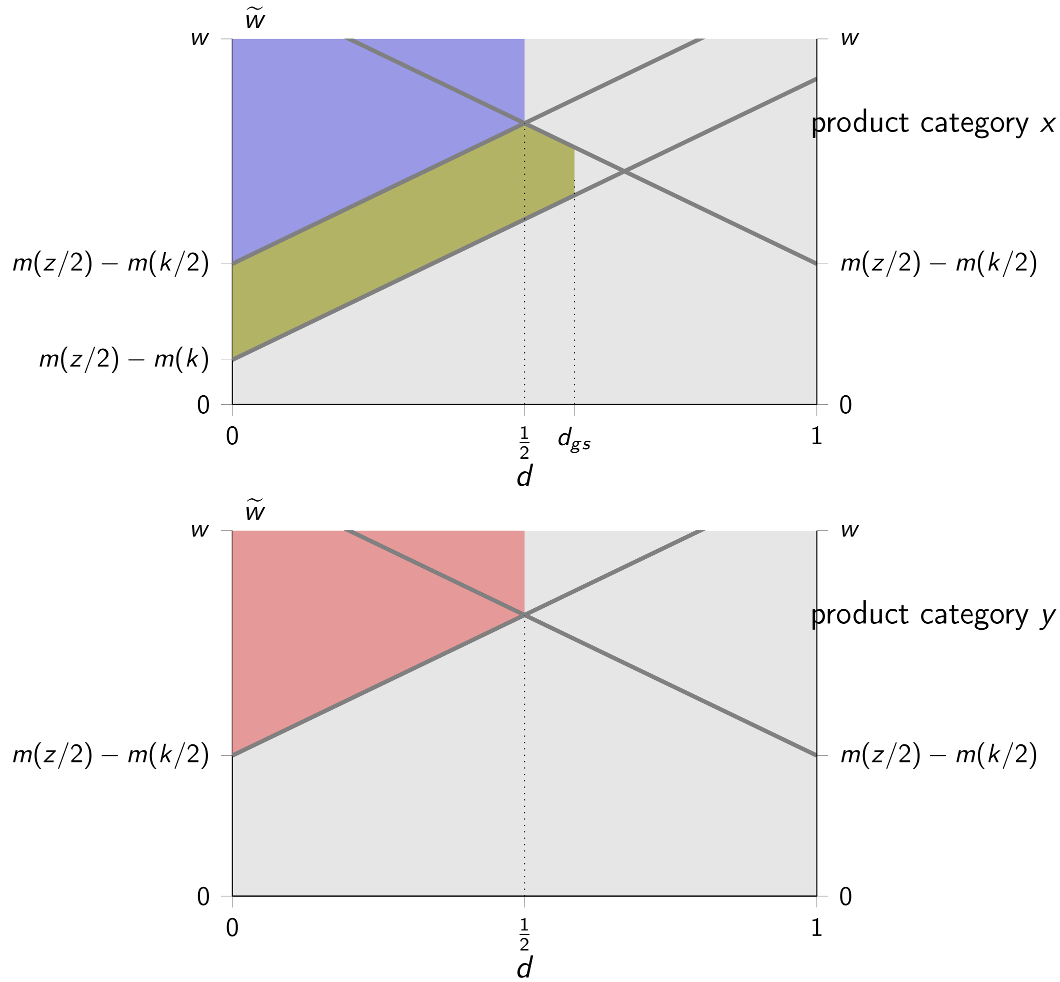
Both cross-partials are negative, confirming that larger stores benefit less from amenity investments. Finally, taking derivatives of (19) with respect to w we get

$$\frac{d^2 \pi_g}{dw^2} < \frac{d^2 \pi_s}{dw^2} < 0$$

which concludes the proof. ■

Proof of Proposition 6: Figure A.1 illustrates the competition case. On the horizontal axis we measure the consumer location d, where $d = 0$ corresponds to local store b_0 and $d = 1$ corresponds to local store b_1. On the vertical axis we measure $\widetilde{w}$, the relative preference for a local store. We assume that d and $\widetilde{w}$ are independently and uniformly distributed: $d \sim U[0, 1]$ and $\widetilde{w} \sim U[0, w]$. Since there are two different product categories, we need to plot one graph per product category, product category x on the top panel and product category y on the

Figure A.1
Store strategy under local competition



bottom panel.

Figure A.1 illustrates the case when both b_0 and b_1 are general stores. Store b_0's demand of product category x is given by the area in blue in the top panel, whereas store b_0's demand of product category y is given by the area in red in the bottom panel. To understand that, notice that store b_0 must beat both store a and store b_1. Beating store a requires

$$m(k/2) + \widetilde{w} - \tau\,\widetilde{d} > m(z/2)$$

whereas beating store b_1 requires

$$m(k/2) - \tau\,\widetilde{d} > m(k/2) - \tau\,(1 - \widetilde{d})$$

This results in the following set of inequalities

$$\begin{aligned} \widetilde{w} &> m(z/2) - m(k/2) + \tau\,\widetilde{d} \\ \widetilde{d} &< \tfrac{1}{2} \end{aligned}$$

which in turn correspond to the areas in blue (top panel) and red (bottom panel).

Given that b_1 chooses to be a general store, how does b_0 change its profits by specializing in product category x? Store b_1's demand from x consumers is now determined by

$$m(k) + \widetilde{w} - \tau\,\widetilde{d} > m(z/2)$$

(beat firm a) and

$$m(k) + \widetilde{w} - \tau\,\widetilde{d} > m(k/2) + \widetilde{w} - \tau\,(1 - \widetilde{d})$$

(beat firm b_1). This simplifies to

$$\begin{aligned} \widetilde{w} &> m(z/2) - m(k) + \tau\,\widetilde{d} \\ \widetilde{d} &< d_{gs} \equiv \tfrac{1}{2} + \big(m(k) - m(k/2)\big)\,/(2\tau) \end{aligned}$$

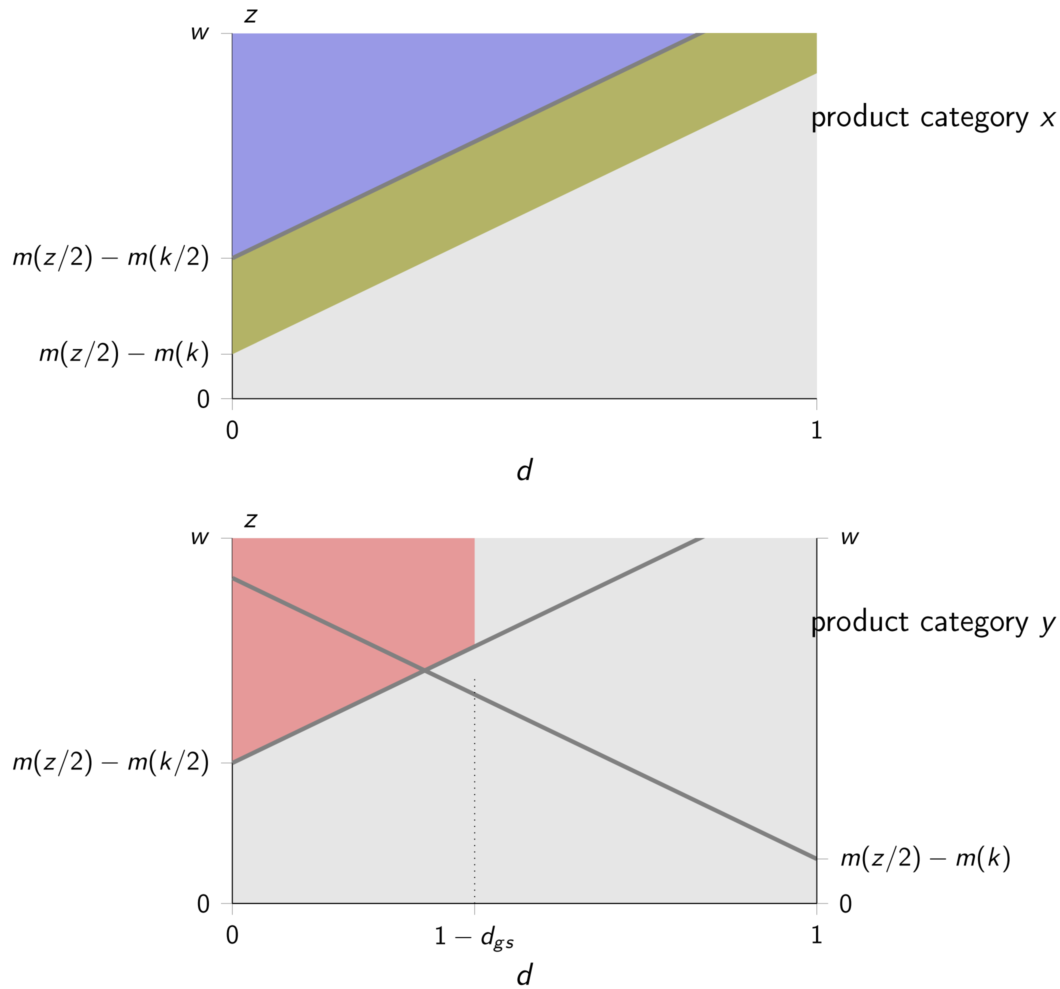
This corresponds to an increase in demand for product category x given by the area in green on the top panel and a loss in demand for product category y given by the area in red on the bottom panel. The green area on the top panel corresponds entirely to consumers who purchased from a when both b_0 and b_1 were general stores and now prefer to buy from b_0, the product category x specialty store. The red area on the bottom panel corresponds to consumers who were interested in store b_0 when it was a general store but are now not interested since it no longer carries any product category y products.

The values of z and k in Figure A.1 were chosen so that the areas in green and red are equal. This implies that, given that store b_1 follows a general-store strategy, store b_0 is indifferent between being a general store and being a specialty store. Suppose now that b_1 chooses to be a y-specialty store. What is the gain for store b_0 from specializing in x? This alternative scenario is described in Figure A.2. In terms of x consumers, the battle is now limited to firms b_0 and a, since firm b_1 is absent from this product category. Demand for firm b_0 is determined by

$$m(k/2) + \widetilde{w} - \tau\,\widetilde{d} > m(z/2)$$

which corresponds to the area in blue. Regarding product category y (bottom panel), we still need to consider both competition by a and competition by b_1. Since b_1 is a product

Figure A.2
Store strategy under local competition



category y specialty store, we now have

$$\begin{aligned} \widetilde{w} &> m(z/2) - m(k) + \tau\,\widetilde{d} \\ \widetilde{d} &< 1 - d_{gs} \equiv \tfrac{1}{2} + \big(m(k/2) - m(k)\big)\,/(2\tau) \end{aligned}$$

which corresponds to the area in red. What happens to firm b_0's profit as it switches from a general store to a product category x specialty store? On the top panel (that is, in terms of x sales), it experiences a profit increase given by the green area. On the bottom panel (that is, in terms of y sales), it experiences a profit loss given by the red area.

Immediate inspection reveals that the green area in the top panel of Figure A.2 is greater than the green area in the top panel of Figure A.1, whereas the red area in the bottom panel of Figure A.2 is lower than the red area in the bottom panel of Figure A.1. This implies that, if firm b_0 is indifferent between being a general store and being a specialty store when its rival is a general store, then it strictly prefers to be a specialty store when its rival is a specialty store. ■

B. Data: Additional Figures and Tables

Figure B.1
General Stores

Neighborhood Shop

Convenience Chain

Supermarket Chain

Source: Authors' photos; https://www.oxxo.com/quienes-somos; www.walmex.mx/nosotros/formatos-de-negocio.html

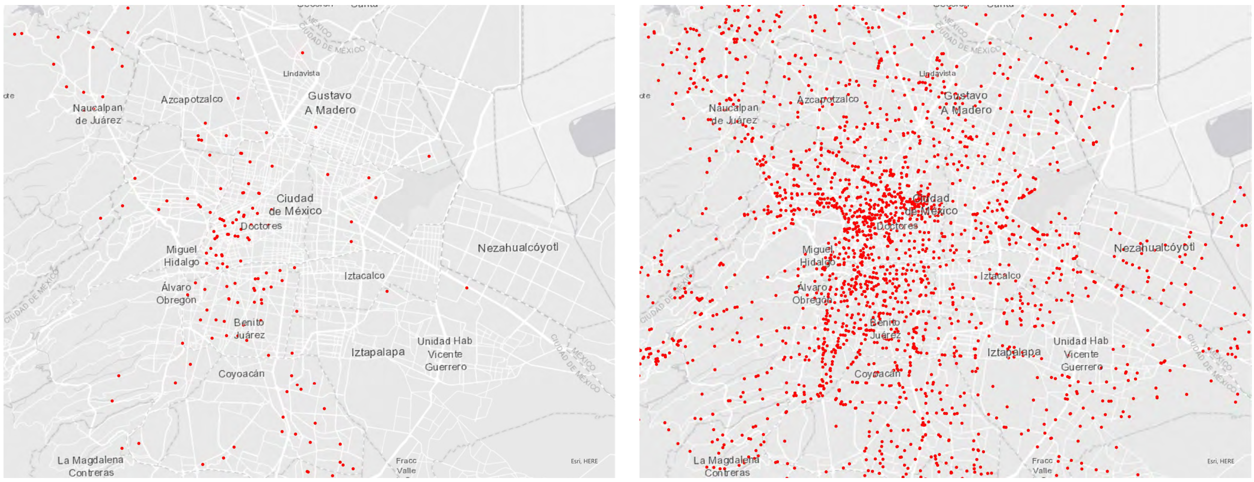
Figure B.2
Specialty Stores

Candy

Dairy and Deli Meat

Bread and Baked Goods

Figure B.3
Convenience Chains in Mexico City



1999 2019

Source: Esri, HERE, DENUE (INEGI), Marco Geostadistico (INEGI)
Note: The maps display the location of convenience chain stores for 1999 and 2019. Locations for 1999 are approximated using the 1999 Economic Census Data. Locations for 2019 are from INEGI's business registry (DENUE).

Table B.1
General and Specialty Stores in 1999

Dependant Variable (log):	Revenue	Profits	Age	Total Employed	Hired Employees	Revenue x Wrkr	Profits x Wrkr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I[Specialist]	-0.002	-0.051***	-0.085***	-0.018***	0.148***	0.007	-0.059***
	(0.013)	(0.011)	(0.008)	(0.004)	(0.006)	(0.012)	(0.011)
Observations	719,866	719,866	719,866	719,866	719,866	717,007	708,800
Census Tract FE	Y	Y	Y	Y	Y	Y	Y

Note: The table displays the differences between general and specialty stores estimated using the following equation: $Y_f = \beta_0 + \beta_1 1[specialty]_f + \delta_{m(f)} + \varepsilon_i$, where Y_f denotes the outcome of interest in logs for firm f, β_1 is the estimated difference between specialty and general stores, and $\delta_{m(f)}$ is the census tract fixed effect. The sample does not include establishments owned by a chain or the government. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Figure B.4

Map of N=554 Stores in Primary Dataset

We report how our survey instruments measured key variables in our within-firm specialization analysis. We report questions as they were originally written in English, though ultimately Spanish translations (completed by a locally-based RA) were used when the survey was launched in the field.

B.1. Baseline Survey

The total number of SKUs. This measure is used to categorize local stores as capacity-constrained (small k).

- Question — How many different PRODUCTS/SERVICES do you sell in your business today? *ENUMERATOR: Enter the total number of products/services that the business currently sells. Be especially careful with retailers who sell many different products, but all in relatively small quantities. Note that different brands/varieties count as different products*
- Response — Single-choice from the following drop-down list: 1= Between 1 and 10 different products; 2= Between 11 and 20 different products; 3= Between 21 and 40 different products; 4= Between 41 and 60 different products; 5= Between 61 and 80 different products; 6= Between 81 and 100 different products; 7= Between 101 and 150 different products; 8= Between 151 and 200 different products; 9= Between 201 and 250 different products; 10= Between 251 and 300 different products; 11= More than 300 different products

Description of the main categories sold. This measure is used in our LLM-based specialization index to identify categories some sold at baseline.

- Question— What do you sell in the primary part of your business? *ENUMERATOR: Please have the respondent explain in detail what they sell so that someone can visualize what this business does and looks like.*
- Response— Open-text box.

Top 3 SKUs sold at the store by volume. This measure is used in our LLM-based specialization index to identify some categories sold at baseline.

- Question—List your 3 MOST POPULAR products or services. These are the three products/services that you sell the most units of. Please enter a description of this product/service and include the unit amount. For example: "sugar, 1 kg bag". *ENUMERATOR: Probe to make sure these 3 products/services are the ones* *for which the respondent sells the MOST UNITS (i.e., highest total volume). They are not necessarily the most expensive offerings of the business, but the products/services that customers buy the most of – i.e., they buy them most often or in the greatest volume.*

- Response 1— Product/Service 1: Open-text box.
- Response 2— Product/Service 2: Open-text box.
- Response 3— Product/Service 3: Open-text box.

B.2. Endline Survey

Description of the main categories sold. This measure is used in our LLM-based specialization index to identify some categories sold at endline.

- Question— What do you sell in the primary part of your business? *ENUMERATOR: Please have the respondent explain in detail what they sell so that someone can visualize what this business does and looks like.*
- Response— Open-text box.

Addition of products. This measure is used on its own as a dependent variable and in our LLM-based specialization index to identify products and categories added to the assortment by endline.

- **Question**—Have you carried out any analysis or actions to introduce a new product or product(s) in your business? By this we mean, have you used your resources (time and money) to add a new product or products(s) to your product assortment? *ENUMERATOR: Make sure the respondent understands what is meant by adding a new product to their business assortment. This does not mean changing the quantities of existing products (i.e. ordering more of a product that was already offered) but adding a completely new product to their business. Possible scenarios are: New brand added-For example, a grocery business was previously selling only Coke and Sprite, now they offer Squirt; New product added within existing brand-For example, a grocery business was previously selling only Coke and Sprite, now they offer Coke Light and Diet Sprite.*
- **Response**— Open-text box. Binary Yes/No on addition coded from text response.
- **Follow-up question**—If you introduced a new product or product(s) in your business, what was the main reason why you decided to introduce it? *ENUMERATOR: DO NOT READ THE OPTIONS. The business owner should verbally*

describe to you their reasons for introducing the product in the business. Please isolate the main reason and classify it into the response categories. If the business owner DOES NOT EXPLICITLY TALK ABOUT MARGINS (option 1) or CUSTOMERS (option 2) or COMPETITORS (option 3) in their reasoning, please classify the reason as 'other'.

- **Follow-up response**–Single-choice from the following drop-down list: 0 = Not applicable since the business owner did not add any new product; 1 = Business owner added new product(s) for margin-related reasons; 2 = Business owner added new product(s) for customer-related reasons (e.g. customers asked for it, or they anticipated customers would demand it); 3 = Business owner added new product(s) for competitor-related reasons (e.g. a business rival offered it so they had to respond, or they wanted to separate themselves from the competition); 4 = Other reason

Removal of products. This measure is used on its own as a dependent variable and in our LLM-based specialization index to identify products and categories removed from the assortment by endline.

- **Question**—Have you carried out any analysis or actions to remove a new product or product(s) in your business? By this we mean, have you removed a product or products(s) from assortment that you were previously offering? *ENUMERATOR: Make sure the respondent understands what is meant by removing a product from their business assortment. This does not mean changing the quantities of existixng products (i.e. ordering less of a product that was already offered) but completely eliminating a product from the business.*
- **Response**— Open-text box. Binary Yes/No on removal coded from text response.
- **Follow-up question**—if you removed a product or product(s) from your assortment, what was the main reason why you decided to stop selling it? *ENUMERATOR: The business owner should verbally describe to you why they removed a product or product(s) from their assortment. Please isolate the main reason and classify it into the response categories.*
- **Follow-up response**–Single-choice from the following drop-down list: 0 = Not applicable since the business owner did not did not remove/stop offering product; 1 = Business owner removed/stopped offering product(s) for margin-related reasons (e.g. these products had a low margin or were too costly to produce); 2 = Business owner removed/stopped offering product(s) for customer-related reasons

(e.g. customers did not show interest in it, or buy it); 3 = Business owner removed/stopped offering product(s) for competitor-related reasons; 4 = Business owner removed/stopped offering product(s) because they personally did not like the product(s); 5 = Other reason

Finally, we provide randomly-sampled examples of responses for product additions and removals in Table B.2.

Addition	**Translation**	**Removal**	**Translation**
El empresario incluyó a su oferta una nueva clase de cupcakes	The businessman added a new type of cupcake to his offerings.	No, el empresario no quitó productos porque está cómodo con su oferta	No, the businessman didn't remove any products because he's comfortable with his offering.
Metió four loko cerveza titanium	Added four loko cerveza titanium	No	No
Los empresarios añadieron Coca-Cola Energy y café, Belight de lichi, Pepsi Black, Levite infusiones y Vitaloe porque sus proveedores sacaron esos nuevos productos.	The businessmen added Coca-Cola Energy and coffee, Belight Lychee, Pepsi Black, Levite infusions and Vitaloe because their suppliers launched those new products.	Quitó de su oferta un garrafón de Epura porque el proveedor fue irresponsable, consomé de Kanbels porque no se vendía, una salsa de tomate Prego porque no se vendía, jamón Swan porque no se vendía y se hechaba a perder, Mirinda porque los clientes no se interesaban por ella.	He removed a Garrafon (container) of Epura because the supplier was irresponsible, Kanbels consommé because it wasn't selling, Prego tomato sauce because it wasn't selling, Swan ham because it wasn't selling and was going to waste, and Mirinda because customers weren't interested in it.
No	No	Si el margen de utilidad era muy poco	Yes, the profit margin was very small
Sí, nuevos Doritos, Coca-Cola Energy, Volt energetizante, Coca-Cola café, Peñafiel manganada	Yes, new Doritos, Coca-Cola Energy, Volt Energy, Coca-Cola Coffee, Peñafiel Manganada	Tostadas artesanales, porque no se vendían	Artisanal tostadas as they were not selling

Table B.2
Exemplar Text Responses at Endline (t=2) from 5 Randomly-Sampled Stores

Table B.3
Summary Statistics of Stores at Baseline (t=1)

	Mean	SD	Census 2019 SCIAN: 461	Census 2019 SCIAN: 465	Census 2019 SCIAN: 722
Monthly Revenue	59,765.78	10,485.70	60,966.01	35,700.70	67,045.12
Monthly Profits	10,449.66	16,571.73	15,029.0	5,326.05	19,130.75
Hired Employees	2.21	2.25	0.52	0.62	2.08
Total Employed	2.51	2.32	1.88	1.84	3.58

Note: The table displays summary statistics on N=554 retail stores in the sample collected at baseline at the business location. All monetary values are expressed in 2018 MXN Pesos. The statistics for the 2019 Census include small and medium-sized establishments (those that do not have a large-firm identifier in the micro data) within the stated 3-digit SCIAN classification and located in Mexico City. Profits are calculated as revenue minus reported expenses (rent, wages, cost of goods, and other costs).

C. Market-level Evidence: Additional Tables

Table C.1
First Stage Estimates: Entry of Large Chains and Specialization in the Traditional Channel

Dependant Variable:	Number of Convenience Chain Stores	Number of Supermarket Chain Stores
	(1)	(2)
Instrument Z_C	0.10***	
	(0.01)	
Instrument Z_S		0.02***
		(0.002)
Observations	160,103	160,103
Year x Mun. FE	Y	Y
Market FE	Y	Y
Mean Ch. Stores	0.3	0.1

Note: The table reports the first-stage estimates for the IV. The dependent variable is the number of convenience chain stores in the census tract (column 1) or the number of supermarket chain stores in the census tract (column 2). The instruments for convenience stores (Z_C) and supermarkets (Z_S) chains are constructed based on equation 3. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.2
Specialization: Additional Outcomes

Dependant Variable:	Convenience Chain Entry				Supermarket Chain Entry			
	% Spec. Value Added	% Spec. Jobs	% Spec. Hours	% Spec. Profits	% Spec. Value Added	% Spec. Jobs	% Spec. Hours	% Spec. Profits
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Chain Stores	6.04***	5.41***	6.16***	6.32***	37.25***	33.27***	37.93***	38.98***
	(0.94)	(0.84)	(0.86)	(0.93)	(6.00)	(6.60)	(6.73)	(6.03)
Observations	158,375	160,101	160,066	157,878	158,375	160,101	160,066	157,878
Year x Mun. FE	Y	Y	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y	Y	Y
Mean Dep. Var	31.3	28.9	26.0	31.4	31.3	28.9	26.0	31.4
Mean Ch. Stores	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0.1
KP-F Statistic	99.7	99.8	99.8	99.7	56.8	55.0	55.1	56.5

Note: The table displays the effect of the entry of a convenience chain store (columns 1-4) or a supermarket (columns 5-8) at the census tract level, estimated based on equation using 2SLS. The % outcomes are the share of the variable for specialized stores over the traditional sector. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.3
Entry of Large Convenience Chains, Food Expenditure, and Trips in the Traditional Channel

Dependant Variable: Share of Consumption and Trips in/to Specialty Stores				
	Consumption ($)		Trips	
	(1)	(2)	(3)	(4)
Number of Conv. Chain Stores	0.42***	0.43***	0.38***	0.40***
	(0.12)	(0.12)	(0.14)	(0.14)
Observations	999,657	999,657	999,657	999,657
Year x Mun. FE	Y	Y	Y	Y
Market FE	Y	Y	Y	Y
HH Controls		Y		Y
Mean Dep. Var	44.4	44.4	47.8	47.8
Mean Ch. Stores	8.6	8.6	8.6	8.6
KP-F Statistic	117.4	117.3	117.4	117.3

Note: The table displays the effect of the entry of a convenience chain store in the market (census tracts within 1 km) estimated through 2SLS based on equation 5 at the household level. The monetary share of consumption is the expenses in specialty stores divided by the expenses in specialty stores and neighborhood shops. The share of trips is the number of days the household visited a specialty store within the week divided by the number of days it visited a specialty store plus the number of days it visited a neighborhood shop. Columns 2 and 4 include household-level controls: income per capita, income, household size, number of adults, age of household head, and monetary expenditures. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.4
Robustness: Specialization in Consumption

Dependant Variable: Share of Consumption and Trips in/to Specialty Stores

	Consumption ($)				Trips			
	1.5 km		2 km		1.5 km		2 km	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Conv.	0.25***	0.25***	0.15***	0.16***	0.25***	0.25***	0.14**	0.15**
Chain Stores	(0.07)	(0.07)	(0.05)	(0.05)	(0.09)	(0.09)	(0.07)	(0.06)
Observations	1,667,381	1,667,381	2,439,339	2,439,339	1,667,381	1,667,381	2,439,339	2,439,339
Year x Mun. FE	Y	Y	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y	Y	Y
HH Controls		Y		Y		Y		Y
Mean Dep. Var	44.6	44.6	44.7	44.7	48.1	48.1	48.3	48.3
Mean Ch. Stores	15.4	15.4	23.9	23.9	15.4	15.4	23.9	23.9
KP-F Statistic	98.3	98.2	108.7	108.7	98.3	98.2	108.7	108.7

Note: The table displays the effect of the entry of a convenience chain store in a given market estimated through 2SLS based on equation 5 at the household level. The monetary share of consumption is the expenses in specialized stores divided by the expenses in specialized stores and neighborhood shops. The share of trips is the number of days the household visited a specialized store within the week divided by the number of days it visited a specialized store plus the number of days it visited a neighborhood shop. Variations across columns denote different definitions of markets (census tracts within 1.5 km, and 2 km) and the inclusion of household-level controls: income per capita, income, household size, number of adults, age of household head, and monetary expenditures. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.5
Robustness: Specialization (Alternative Market Size Definitions)

Dependant Variable: % Specialty (N)

	Convenience Chain Entry					Supermarket Chain Entry				
	AGEB	0.5 km	1 km	1.5 km	2 km	AGEB	0.5 km	1 km	1.5 km	2 km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number of	4.73***	1.11***	0.53***	0.31***	0.20***	29.11***	9.80***	6.51***	3.95***	2.59***
Chain Stores	(0.73)	(0.21)	(0.10)	(0.07)	(0.05)	(5.79)	(2.77)	(2.17)	(1.45)	(0.95)
Observations	160,103	160,242	160,300	160,321	160,324	160,103	160,242	160,300	160,321	160,324
Year x Mun. FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mean Dep. Var	27.9	22.3	21.6	21.3	21.2	27.9	22.3	21.6	21.3	21.2
Mean Ch. Stores	0.3	1.8	3.9	6.8	10.3	0.1	0.4	1.0	1.7	2.5
KP-F Statistic	99.8	135.8	149.4	154.6	152.7	55.0	29.6	14.5	12.3	15.9

Note: The table displays the effect of the entry of a convenience chain store (columns 1-5) or a supermarket (columns 6-10) in a given market on the specialization of the traditional channel, estimated based on equation 5 using 2SLS. The outcome is the share of specialized stores over the traditional sector. Variations across columns denote different definitions of markets (census tract, census tract within 0.5 km, 1 km, 1.5 km, and 2 km). Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.6

Robustness: Market Specialization (3rd Degree Neighboring Cities in Instrument)

	Convenience Chain Entry				Supermarket Chain Entry			
Dependant Variable:	% Specialized (N)	Generalist Stores (N)	Specialized Stores (N)	% Spec. Revenue	% Specialized (N)	Generalist Stores (N)	Specialized Stores (N)	% Spec. Revenue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Chain Stores	5.05***	-2.66***	1.17***	6.05***	29.73***	-15.65***	6.89***	35.58***
	(0.70)	(0.38)	(0.44)	(0.79)	(5.31)	(2.10)	(2.46)	(5.49)
Observations	160,103	160,103	160,103	160,038	160,103	160,103	160,103	160,038
Year x Mun. FE	Y	Y	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y	Y	Y
Mean Dep. Var	27.9	12.8	8.0	32.1	27.9	12.8	8.0	32.1
Mean Ch. Stores	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0.1
KP-F Statistic	140.5	140.5	140.5	140.6	72.7	72.7	72.7	72.7

Note: The table displays the effect of the entry of a convenience chain store (columns 1-4) or a supermarket (columns 5-8) at the census tract level, estimated based on equation 5 but using third-degree neighboring cities instead of second-degree neighboring cities to construct the instrument. The % outcomes are the share of the variable for specialized stores over the traditional sector. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.7

Robustness: Specialization (Alternative Definitions of Specialized Stores)

Dependant Variable: % Specialty (N)						
	Convenience Chain Entry			Supermarket Chain Entry		
	461	4611	4612	461	4611	4612
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Chain Stores	4.73***	5.32***	3.45***	29.11***	32.70***	21.46***
	(0.73)	(0.79)	(0.49)	(5.79)	(6.24)	(3.72)
Observations	160,103	158,480	157,976	160,103	158,480	157,976
Year x Mun. FE	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y
Mean Dep. Var	27.9	22.3	9.7	27.9	22.3	9.7
Mean Ch. Stores	0.3	0.3	0.3	0.1	0.1	0.1
KP-F Statistic	99.8	98.8	96.0	55.0	54.8	50.7

Note: The table displays the effect of the entry of a convenience chain store (columns 1-3) or a supermarket (columns 4-6) in the census tract on the specialization of the traditional channel, estimated based on equation 5 using 2SLS. The outcome is the share of specialized stores in the traditional sector, and it always includes all general stores in the denominator. The specialty stores included in both the numerator and the denominator vary across specifications: i) all of them, 3-digit classification 461 (establishments within retail sale of groceries, food, beverages, ice, and tobacco), ii) 4-digit classification 4611 (establishments within retail sale of groceries and food), and iii) 4-digit classification 4612 (establishments within beverages, ice, and tobacco). Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table C.8

OLS Estimates: Entry of Large Chains and Specialization in the Traditional Channel

	Convenience Chain Entry				Supermarket Chain Entry			
Dependant Variable:	% Specialty (N)	General Stores (N)	Specialty Stores (N)	% Spec. Revenue	% Specialty (N)	General Stores (N)	Specialty Stores (N)	% Spec. Revenue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Chain Stores	7.76***	0.41**	4.65***	10.00***	13.88***	1.55***	11.92***	17.72***
	(0.28)	(0.20)	(0.67)	(0.39)	(0.45)	(0.39)	(4.30)	(0.52)
Observations	164,584	164,584	164,584	164,516	164,584	164,584	164,584	164,516
Year x Mun. FE								
Market FE								
Mean Dep. Var	27.9	12.8	8.0	32.1	27.9	12.8	8.0	32.1
Mean Ch. Stores	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0.1

Note: The table displays the effect of the entry of a convenience chain store (columns 1-4) or a supermarket (columns 5-8) at the census tract level, estimated based on equation 5, but without instrumenting for the number of convenience chain or supermarket chain stores and without market and municipality-year fixed effects. The % Specialty (N) is the number of specialty stores divided by the total general and specialty stores in the traditional sector in the census tract. The % Spec. Revenue is the total revenue of specialty stores divided by the total revenue of general and specialty stores in the traditional sector in the census tract. Standard errors in parentheses are clustered at the municipality level. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

D. Within-Firm Evidence: Additional Figures and Tables

First, we report the distribution of time elapsed between the baseline (t=1) and endline (t=2) surveys. The average gap is 19.7 months. Because this interval is short relative to the five-year spacing of census waves, the scope for market-specific time-varying shocks such as gentrification to drive chain entry and within-firm specialization is limited.

Figure D.1
Time Between Baseline and Endline Surveys

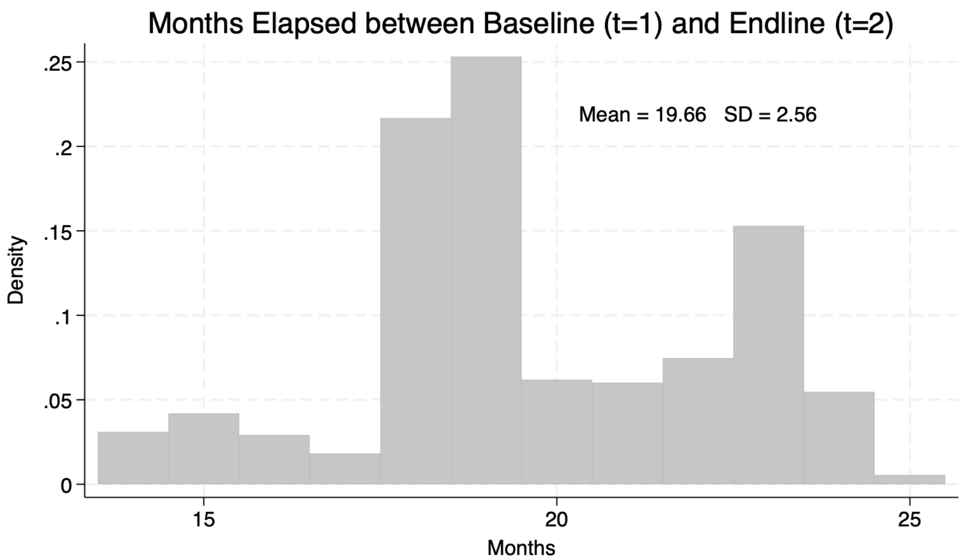


Table D.1

Robustness: Within-Firm Assortment Changes (Alternative Trade Area Definitions: 2 km)

	Assortment Changes: Competition/Customers			Assortment Changes: Placebo Reasons		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.11***	0.12**	0.15***	0.02	-0.04	-0.01
	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)
Observations	550	550	550	550	550	550
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.17	0.33	0.42	0.22	0.26	0.41

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 2 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes, respectively: assortment removal due to competition or customers, addition due to competition or customers, change due to competition or customers, removal due to other (placebo) reasons, addition due to other (placebo) reasons, and change due to other (placebo) reasons. The indicated regressions include store characteristic controls for precision (monthly profits, total employees, number of weekly customers, tax registration status, and an index of business practices). Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.2

Robustness: Within-Firm Assortment Changes (Alternative Trade Area Definition: 0.5 km)

	Assortment Changes: Competition/Customers			Assortment Changes: Placebo Reasons		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.11**	0.09	0.11*	-0.01	-0.04	-0.03
	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)
Observations	385	385	385	385	385	385
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.21	0.36	0.48	0.24	0.27	0.43

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 0.5 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes, respectively: assortment removal due to competition or customers, addition due to competition or customers, change due to competition or customers, removal due to other (placebo) reasons, addition due to other (placebo) reasons, and change due to other (placebo) reasons. The indicated regressions include store characteristic controls for precision (monthly profits, total employees, number of weekly customers, tax registration status, and an index of business practices). Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.3
Robustness: Within-Firm Assortment Changes (No Outlier Trimming)

	Assortment Changes: Competition/Customers			Assortment Changes: Placebo Reasons		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.14***	0.14***	0.18***	0.01	-0.07	-0.04
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Observations	391	391	391	391	391	391
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.17	0.33	0.42	0.22	0.26	0.41

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes, respectively: assortment removal due to competition or customers, addition due to competition or customers, change due to competition or customers, removal due to other (placebo) reasons, addition due to other (placebo) reasons, and change due to other (placebo) reasons. The indicated regressions include store characteristic controls for precision (monthly profits, total employees, number of weekly customers, tax registration status, and an index of business practices). Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.4
Robustness: Within-Firm Assortment Changes (5% Outlier Trimming)

	Assortment Changes: Competition/Customers			Assortment Changes: Placebo Reasons		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.14***	0.15***	0.19***	0.00	-0.07	-0.05
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Observations	384	384	384	384	384	384
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.16	0.33	0.42	0.22	0.26	0.42

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes, respectively: assortment removal due to competition or customers, addition due to competition or customers, change due to competition or customers, removal due to other (placebo) reasons, addition due to other (placebo) reasons, and change due to other (placebo) reasons. The indicated regressions include store characteristic controls for precision (monthly profits, total employees, number of weekly customers, tax registration status, and an index of business practices). Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.5
Robustness: Within-Firm Assortment Changes (Food Retailers Only)

	Assortment Changes: Competition/Customers			Assortment Changes: Placebo Reasons		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.15**	0.20***	0.24***	0.05	-0.04	0.02
	(0.06)	(0.07)	(0.08)	(0.06)	(0.06)	(0.08)
Observations	264	264	264	264	264	264
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.15	0.27	0.37	0.17	0.23	0.35

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes, respectively: assortment removal due to competition or customers, addition due to competition or customers, change due to competition or customers, removal due to other (placebo) reasons, addition due to other (placebo) reasons, and change due to other (placebo) reasons. The indicated regressions include store characteristic controls for precision (monthly profits, total employees, number of weekly customers, tax registration status, and an index of business practices). Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.6
Robustness: Within-Firm Assortment Changes by Capacity Constraints (Alternative Trade Area Definitions: 2 km)

	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.15***	0.17***	0.21***	-0.01	0.03	0.05
	(0.04)	(0.06)	(0.06)	(0.09)	(0.09)	(0.10)
Observations	344	344	344	206	206	206
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.10	0.33	0.40	0.34	0.31	0.47

Note: This table displays the assortment changes within stores resulting from the entry of large chains within a 2 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to competition or customers for more capacity-constrained stores, addition due to competition or customers for more capacity-constrained stores, change due to competition or customers for more capacity-constrained stores, removal due to competition or customers for less capacity-constrained stores, addition due to competition or customers for less capacity-constrained stores, and change due to competition or customers for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.7
Robustness: Within-Firm Assortment Changes by Capacity Constraints (Alternative Trade Area Definition: 0.5 km)

	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.15**	0.17**	0.20***	0.01	-0.01	-0.03
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.09)
Observations	247	247	247	138	138	138
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.15	0.38	0.46	0.34	0.31	0.51

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 0.5 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to competition or customers for more capacity-constrained stores, addition due to competition or customers for more capacity-constrained stores, change due to competition or customers for more capacity-constrained stores, removal due to competition or customers for less capacity-constrained stores, addition due to competition or customers for less capacity-constrained stores, and change due to competition or customers for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.8
Robustness: Within-Firm Assortment Changes by Capacity Constraints (No Outlier Trimming)

	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.19***	0.24***	0.28***	-0.02	0.01	0.04
	(0.05)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)
Observations	240	240	240	151	151	151
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.10	0.33	0.40	0.34	0.31	0.47

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to competition or customers for more capacity-constrained stores, addition due to competition or customers for more capacity-constrained stores, change due to competition or customers for more capacity-constrained stores, removal due to competition or customers for less capacity-constrained stores, addition due to competition or customers for less capacity-constrained stores, and change due to competition or customers for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.9

Robustness: Within-Firm Assortment Changes by Capacity Constraints (5% Outlier Trimming)

	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.20***	0.24***	0.28***	-0.01	0.01	0.05
	(0.05)	(0.07)	(0.07)	(0.10)	(0.10)	(0.11)
Observations	239	239	239	145	145	145
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.10	0.33	0.40	0.33	0.33	0.47

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to competition or customers for more capacity-constrained stores, addition due to competition or customers for more capacity-constrained stores, change due to competition or customers for more capacity-constrained stores, removal due to competition or customers for less capacity-constrained stores, addition due to competition or customers for less capacity-constrained stores, and change due to competition or customers for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.10

Robustness: Within-Firm Assortment Changes by Capacity Constraints (Food Retailers Only)

	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
Dependant Variable:	Removed	Added	Changed	Removed	Added	Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.26***	0.33***	0.38***	-0.09	0.06	0.03
	(0.06)	(0.09)	(0.09)	(0.13)	(0.12)	(0.14)
Observations	152	152	152	112	112	112
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.06	0.28	0.33	0.38	0.25	0.44

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to competition or customers for more capacity-constrained stores, addition due to competition or customers for more capacity-constrained stores, change due to competition or customers for more capacity-constrained stores, removal due to competition or customers for less capacity-constrained stores, addition due to competition or customers for less capacity-constrained stores, and change due to competition or customers for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.11
Robustness: Within-Firm Placebo Assortment Changes by Capacity Constraints

Dependant Variable:	More Capacity-Constrained (small k)			Less Capacity-Constrained (large k)		
	Placebo: Removed	Placebo: Added	Placebo: Changed	Placebo: Removed	Placebo: Added	Placebo: Changed
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	-0.03	-0.07	-0.09	0.10	-0.01	0.14
	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.09)
Observations	239	239	239	149	149	149
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.25	0.30	0.48	0.12	0.16	0.22

Note: The table displays the assortment changes within stores resulting from the entry of large chains within a 1 km radius, estimating equation 6 using a linear probability model with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. Variation across columns denote diverse outcomes and subgroups, respectively: assortment removal due to placebo reasons for more capacity-constrained stores, addition due to placebo reasons for more capacity-constrained stores, change due to placebo reasons for more capacity-constrained stores, removal due to placebo reasons for less capacity-constrained stores, addition due to placebo reasons for less capacity-constrained stores, and change due to placebo reasons for less capacity-constrained stores. Robust standard errors are in parentheses. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.12
Robustness: Within-Firm Specialization by Capacity Constraints (Alternative Trade Area Definition: 2 km)

Dependant Variable:	Specialization Index			Raw Depth Ratio (Std.)		
	Full	Small k	Large k	Full	Small k	Large k
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.18*	0.24**	0.02	0.15*	0.23**	-0.05
	(0.09)	(0.11)	(0.20)	(0.09)	(0.10)	(0.20)
Observations	549	344	205	549	344	205
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	-0.14	-0.16	-0.07	-0.12	-0.16	-0.01

Note: The table displays the specialization within stores resulting from the entry of large chains within a 2 km radius, estimated using OLS regressions with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. The dependent variables are the two specialization measures described in Section 6.1: the LLM-based index and raw assortment depth ratio. Variation across columns denote different subgroups, respectively: the full sample, more capacity-constrained stores, and less capacity-constrained stores. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.13

Robustness: Within-Firm Specialization by Capacity Constraints (Alternative Trade Area Definition: 0.5 km)

	Specialization Index			Raw Depth Ratio (Std.)		
Dependant Variable:	Full	Small k	Large k	Full	Small k	Large k
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.21*	0.36**	0.02	0.20*	0.39**	-0.04
	(0.11)	(0.15)	(0.16)	(0.12)	(0.17)	(0.17)
Observations	384	247	137	384	247	137
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	-0.06	-0.06	-0.04	-0.06	-0.07	-0.03

Note: The table displays the specialization within stores resulting from the entry of large chains within a 0.5 km radius, estimated using OLS regressions with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. The dependent variables are the two specialization measures described in Section 6.1: the LLM-based index and raw assortment depth ratio. Variation across columns denote different subgroups, respectively: the full sample, more capacity-constrained stores, and less capacity-constrained stores. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.14

Robustness: Within-Firm Specialization by Capacity Constraints (No Outlier Trimming)

	Specialization Index			Raw Depth Ratio (Std.)		
Dependant Variable:	Full	Small k	Large k	Full	Small k	Large k
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.21**	0.28**	0.07	0.19*	0.28**	0.01
	(0.10)	(0.12)	(0.22)	(0.10)	(0.12)	(0.22)
Observations	391	240	151	391	240	151
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	-0.14	-0.16	-0.07	-0.12	-0.16	-0.01

Note: The table displays the specialization within stores resulting from the entry of large chains within a 1 km radius, estimated using OLS regressions with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. The dependent variables are the two specialization measures described in Section 6.1: the LLM-based index and raw assortment depth ratio. Variation across columns denote different subgroups, respectively: the full sample, more capacity-constrained stores, and less capacity-constrained stores. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.15
Robustness: Within-Firm Specialization by Capacity Constraints (5% Outlier Trimming)

	Specialization Index			Raw Depth Ratio (Std.)		
Dependant Variable:	Full	Small k	Large k	Full	Small k	Large k
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.21**	0.28**	0.06	0.20*	0.29**	-0.01
	(0.11)	(0.12)	(0.23)	(0.10)	(0.12)	(0.23)
Observations	384	239	145	384	239	145
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	-0.13	-0.16	-0.04	-0.11	-0.16	0.03

Note: The table displays the specialization within stores resulting from the entry of large chains within a 1 km radius, estimated using OLS regressions with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. The dependent variables are the two specialization measures described in Section 6.1: the LLM-based index and raw assortment depth ratio. Variation across columns denote different subgroups, respectively: the full sample, more capacity-constrained stores, and less capacity-constrained stores. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.

Table D.16
Robustness: Within-Firm Specialization by Capacity Constraints (Food Retailers Only)

	Specialization Index			Raw Depth Ratio (Std.)		
Dependant Variable:	Full	Small k	Large k	Full	Small k	Large k
	(1)	(2)	(3)	(4)	(5)	(6)
Large Chain Entered	0.10	0.30**	-0.24	0.12	0.34**	-0.26
	(0.13)	(0.14)	(0.27)	(0.14)	(0.13)	(0.31)
Observations	264	152	112	264	152	112
Store Controls	Y	Y	Y	Y	Y	Y
Mean Dep. Var. Control	0.07	-0.06	0.37	0.03	-0.13	0.39

Note: The table displays the specialization within stores resulting from the entry of large chains within a 1 km radius, estimated using OLS regressions with observations at the store level. Our control group consists of stores who did not experience the entry of large chains within 2 km, for consistency with our main specification. The dependent variables are the two specialization measures described in Section 6.1: the LLM-based index and raw assortment depth ratio. Variation across columns denote different subgroups, respectively: the full sample, more capacity-constrained stores, and less capacity-constrained stores. The stars next to the estimate, *, **, ***, represent statistical significance at the .10, .05, and .01 level, respectively.