Consumer Search and Firm Location: Theory and Evidence from the Garment Sector in Uganda

Anna Vitali*

September 25, 2025

Abstract

This paper examines how consumer information frictions shape firm location decisions in low-income cities. Using an original survey and transaction data from garment firms and their customers in Kampala, Uganda, I show that product search is costly and that consumers in dense areas visit more businesses and form better matches. Guided by these patterns, I develop a quantitative spatial model of consumer search and firm location, in which agglomeration reduces the cost of acquiring information. I estimate the model using a uniquely rich dataset combining prices, product quality, production, and search behavior. I find that information frictions are a key driver of firm agglomeration, comparable in magnitude to widely studied production externalities. I use the model to evaluate policies aimed at decongesting city centers, a priority for many low-income country governments seeking to preserve urban livability. Counterfactuals show that e-commerce reduces agglomeration and shifts demand toward higher-quality firms, while evicting businesses without reducing search costs disproportionately harms them. Ignoring information frictions leads to biased predictions, understating both the welfare gains from e-commerce and the costs of eviction policies.

Key words: Firm location, consumer search, information frictions, low-income cities
JEL codes: R12, R30, D83, L11, O18

*New York University. Email: annavitali@nyu.edu. I am indebted to Imran Rasul, Fabien Postel-Vinay, Costas Arkolakis and Oriana Bandiera for the guidance and support throughout this project. I also thank Treb Allen, Orazio Attanasio, Vittorio Bassi, Robin Burgess, Joao Granja, Jonas Hjort, Aureo de Paula, Suphanit Piyapromdee, Martin Rotemberg, John Spray, Bob Staiger, Michela Tincani, Gabriel Ulyssea, Dan Waldinger, Nathan Zorzi, and seminar participants at various institutions for suggestions that have substantially improved this paper. I gratefully acknowledge financial support from the Private Enterprise Development in Low Income Countries initiative and the International Growth Centre.

1 Introduction

The global urban population is projected to grow by 2.5 billion by 2050, with nearly 90% of this growth concentrated in Africa and Asia (UN, 2018). Unlike the historical experience of high-income countries, urbanization in low-income settings is taking place under very different conditions—most notably, at much lower levels of GDP per capita (Bryan et al., 2020). As a result, urban residents in developing countries face a unique set of constraints that shape where they live, work, and shop (Bryan et al., 2025). Understanding how these constraints influence the internal structure of cities is essential for designing urban policies that can accommodate rapid population growth while supporting economic development.

This paper examines one such constraint: consumers' limited information about the types of products firms sell. In many low-income cities, internet penetration remains low.[1] Without reliable directories or online search tools, consumers often do not know which firms sell the products they need and must instead visit them in person to acquire this information—an activity that is costly and time-consuming (Lagakos, 2016; Startz, 2021). These search costs are particularly high when firms are spatially dispersed, leading consumers to favor dense areas, where information is cheaper to acquire. This, in turn, creates an incentive for firms to agglomerate in order to attract foot traffic and increase their visibility.[2] This mechanism is absent from quantitative urban models, which emphasize production externalities as the primary driver of agglomeration (Redding, 2023). However, in low-income settings, where most firms are single-person operations and production and sales occur in the same location (Bandiera et al., 2022), visibility to customers is as critical as minimizing production costs.

In this paper, I examine the role of consumer information frictions in shaping the location choices of garment firms in Kampala, Uganda. I combine novel data from a representative sample of firms and their customers to document new facts about consumer search behavior in a low-income urban context. Building on this evidence, I develop and estimate an equilibrium model of consumer search and firm location that incorporates both information frictions and production externalities as sources of agglomeration. The unique availability of data on prices, quality, production, customer origins, and search behavior allows me to

[1]In Sub-Saharan Africa and South Asia, only 31.2% and 37.9% of formal firms with 5 to 19 employees, respectively, report having a website, compared to 89.2% in North America and 69.4% in Europe and Central Asia (World Bank Enterprise Surveys). This gap is likely even wider for the smaller, informal businesses that dominate low-income economies. See Figure A1 for a detailed visualization of the relationship between internet penetration and GDP per capita.

[2]A longstanding theoretical literature has examined how consumer search behavior shapes firms' location decisions in the presence of information frictions (e.g., Stahl (1982), Wolinsky (1983), Dudey (1990), Fischer and Harrington Jr (1996), Konishi (2005)). However, to my knowledge, there is no empirical work quantifying the contribution of this mechanism to observed firm agglomeration within cities.

separately identify these mechanisms. I use the model for two purposes. First, I quantify the importance of information frictions as a driver of firm agglomeration, and analyze their implications for firm profits and consumer welfare. Second, I evaluate the welfare effects of two policies aimed at decongesting city centers in developing countries: e-commerce and firm evictions. Comparing counterfactuals with and without information frictions, I show that omitting them understates the welfare gains from e-commerce, the costs of evictions, and their distributional consequences for low- versus high-quality firms.

The study relies on original data from a new survey of 600 garment firms and their customers in Kampala. Firms were randomly selected from a census of over 2,400 establishments. The dataset includes four components: (i) a firm survey capturing the production process and potential production externalities; (ii) detailed transaction records—rarely available for informal firms—that include prices, quantities, customer type (final vs. retailer), and place of origin, and form the basis for demand estimation; (iii) a customer survey, based on a random sample of 600 buyers from the transaction records, designed to elicit search behavior; and (iv) a mystery shopper exercise, in which interviewers posed as clients and purchased an identical garment from each firm, allowing for objective measurement of product quality and prices.

The data reveal three key patterns in consumer search behavior. First, search is costly: 91% of transactions occur in person, and consumers spend nearly 9% of the transaction value on transport, with trips to the dense city center—the *core*—costing about three times more than trips within the *periphery*. Second, despite these costs, consumers report preferring to search in the core because of its high firm density, which makes it easier to identify businesses that sell the desired products. Consistent with this, consumers in the core visit 31% more firms before purchasing, in line with density reducing within-location search costs. Third, consumers searching in the core find higher quality products and form longer-lasting relationships, indicating that more intensive search produces better matches. As a result, firms in the core earn twice the daily revenue of firms in the periphery, largely through bulk purchases by retailers who benefit from economies of scale in transport. These gains, however, are partly offset by higher commuting, wage and rental costs faced by firm owners in this area.

I build a model that accounts for these patterns by incorporating information frictions, transport costs, and heterogeneous consumers (final consumers vs. retailers) into a discrete choice model of demand. Prior to searching, consumers do not observe which firms sell the variety of products they seek—for example, because firms lack an online presence, making it difficult to access information on the products they offer. To observe these varieties, consumers must incur a transport cost that depends on the distance between their location and the firm. Once in a location, consumers search firms sequentially and face a marginal cost of visiting each additional firm that is decreasing in firm density. Since high-density locations tend to be farther from residential areas, consumers face a trade-off between higher transport costs *to* dense areas and lower search costs *within* them. This trade-off is less severe for retailers purchasing in bulk, who benefit from economies of scale in transport, and are therefore more likely to buy from spatially concentrated firms.

Firms sell horizontally differentiated products and are heterogeneous in terms of quality, location preferences and commuting distance of the owner. They choose their location simultaneously in a static entry game. Once in a location, firms decide on the optimal combination of land and labor to employ in production and compete by setting prices. The presence of an additional business in the location affects a firm's variable profits—and thus its location choice—through three channels: (i) it attracts a larger customer base to the location (the *market-size* effect); (ii) it intensifies competition within the location (the *market-share* effect); (iii) it reduces the marginal cost of production (the *production externality*). Importantly, the trade-off between market-size and market-share effects differs for high- and low-quality firms. While all firms benefit equally from the larger customer base associated with lower search costs, high-quality firms are less exposed to business stealing in dense areas, as the higher product quality shields them from direct competition. As a result, they are disproportionately more likely to locate in areas with a high concentration of firms.

I use the newly collected data to estimate the model. First, I combine firm transaction records with quality data from the mystery shoppers exercise to estimate demand. A key feature of the data is that, for each transaction, I observe prices, the customer's origin location, and whether the buyer is a final consumer or a retailer. This allows me to separately identify elasticities with respect to prices, distance, and firm density. In particular, I identify consumption externalities arising from information frictions by comparing the share of customers purchasing from firms that are equally distant, offer identical prices and observable characteristics, but are located in areas with different firm density. By holding the number of firms constant, this variation isolates the effect of search costs from the utility gains associated with greater product variety. Price data—typically unavailable in firm-level datasets used to estimate quantitative spatial models—is crucial for disentangling consumption from production externalities. This is because they allow controlling for the presence of production externalities that, by lowering production costs and thus prices, may attract both customers and firms to high-density locations.

Using the estimated parameters, I examine how equilibrium outcomes change in the absence of information frictions. To do so, I re-compute the equilibrium assuming that consumers can observe all product varieties at no search cost. Eliminating these frictions induces 18.8% of firms to relocate outside the core, an effect comparable in magnitude to that of production externalities, which account for *up to* 25.9% of firm agglomeration. Because most relocating firms are high-quality, this shift reduces the share of sales concentrated in the core by 39%. Overall, consumer welfare rises by 9.5%, reflecting both costless search and lower average transport costs from firms relocating to the periphery. By inducing more customers to search rather than opt for the outside option, demand expands, leading to a 3.8% increase in firm profits.

Decongesting city centres has become a priority for many governments in developing countries, driven by mounting pressures from traffic congestion, environmental concerns, and deteriorating urban livability (World Bank, 2021; Gechter and Kala, 2025). I use the model to assess the effects of two policies that the Kampala administration is considering to reduce congestion in the core: (i) the introduction of an e-commerce platform where informal businesses can sell their products, and (ii) the eviction of a subset of firms from the core. The former addresses the root cause of the inefficiency—information frictions—while the latter tackles only its symptom—agglomeration—without resolving the underlying cause.

Introducing an e-commerce platform reduces the number of firms operating in the core by 32%. This is a result of the platform enabling customers to observe the variety of products sold by all firms in the market and imposing a flat delivery fee across the city, which makes the core's geographical centrality irrelevant. Removing information frictions expands consumers' choice set, shifting demand toward high-quality firms that were previously less visible. As a result, the policy primarily benefits high-quality firms, whose profit gains are 29% larger than those of low-quality firms.

Policies that solely relocate firms without addressing information frictions can have unintended consequences. Specifically, evicting firms from the core to mimic the spatial distribution observed in the e-commerce counterfactual unambiguously lowers profits. High-quality firms suffer the largest losses (11.9% vs. 7.7% for low-quality firms), as relegating them to the periphery reduces their visibility to customers and prevents them from exploiting their competitive advantage.

I re-estimate the effects of e-commerce and evictions under a model specification without information frictions to assess the consequences of omitting this feature when evaluating urban policies. In this alternative model, e-commerce has negligible effects on firm location, since the policy does not change firms' incentive to agglomerate to achieve visibility. As a result, the estimated impacts on firm profits and consumer welfare are much smaller (1.1% vs. 8.5%

and 3.3% vs. 30.4%, respectively). Similarly, the model underestimates the total welfare losses from evictions (1.2% vs. 2.3% in the baseline model). Crucially, the counterfactual fails to capture the heterogeneous effects across firms, predicting similar e-commerce impacts for high- and low-quality firms and, if anything, larger losses from evictions for low-quality businesses.

This paper contributes to a growing literature on the structure of cities in developing countries. One strand of this work compares the spatial organization of cities across income levels, documenting how patterns of density, earnings, and built-up area differ between developed and developing contexts (Chauvin et al., 2017; Jedwab et al., 2021; Deffebach et al., 2025; Rosenthal-Kay, 2025). Another strand applies quantitative urban models to evaluate the effects of specific policies, such as investments in transport infrastructure (Balboni et al., 2020; Zárate, 2020; Michaels et al., 2021; Tsivanidis, 2022; Khanna et al., 2023; Kreindler et al., 2023), residential infrastructure (Michaels et al., 2021), and slum upgrading programs (Harari et al., 2018; Gechter and Tsivanidis, 2023; Gertler et al., 2024). While this work has documented spatial patterns and estimated elasticities specific to low-income countries, few studies have explicitly examined the market frictions that distinguish these settings from the assumptions embedded in canonical spatial models, or quantified how these frictions alter policy predictions. This paper takes a step in that direction by focusing on one such friction: consumers' limited access to information about the variety of products sold by different firms. I show that this friction creates additional demand-side incentives for firm agglomeration, which may help explain the higher concentration of employment and built-up density observed in the urban cores of developing-country cities (Deffebach et al., 2025; Rosenthal-Kay, 2025).

More broadly, the paper contributes to the literature using quantitative spatial models to study the organization of economic activity within cities (Ahlfeldt et al., 2015; Allen et al., 2015; Monte et al., 2018; Dingel and Tintelnot, 2020; Owens et al., 2020; Redding, 2023). While this literature has primarily emphasized the role of *production externalities*—such as knowledge spillovers, input-output linkages, and labor market pooling—as key drivers of firm co-location (Duranton and Puga, 2004; Rosenthal and Strange, 2004), demand elasticities in these models are typically assumed to be orthogonal or increasing in firm density, thus capturing only business-stealing effects. This paper expands the framework by introducing *consumption externalities* that arise from consumers' search behavior, which can instead reduce competitive pressure and increase the attractiveness of dense firm clusters.[3] I show

[3]To model consumers' behavior, the paper draws on the industrial organization literature on consumer search with limited information about product characteristics (Hortaçsu and Syverson, 2004; Hong and Shum, 2006; Santos et al., 2012). It is especially close to Murry and Zhou (2020) and Moraga-González

that omitting this channel can lead to biased predictions about the effects of key place-based policies. In doing so, this paper also contributes to a recent strand of urban research exploring the implications of traveling to consume for the spatial distribution of economic activity (Couture, 2016; Agarwal et al., 2017; Miyauchi et al., 2021; Oh and Seo, 2022).

This paper is also part of a growing literature on the role of information frictions in shaping trade flows (Allen, 2014; Startz, 2021; Bergquist et al., 2021), migration decisions (Bryan et al., 2014; Baseler, 2023), and the sorting of firms into profitable locations (Pelnik, 2024) in developing countries. Existing evidence shows that such frictions distort buyer-seller matching, generating excess price dispersion (Jensen, 2007; Aker, 2010; Goyal, 2010; Bergquist et al., 2021) and sustaining the survival of low-quality, low-productivity firms (Atkin et al., 2017; Jensen and Miller, 2018; Pelnik, 2024). I contribute to this literature by using demand-driven agglomeration patterns to infer the magnitude of a specific information friction: consumers' limited knowledge about firm products. I then assess the welfare costs of these frictions, accounting not only for their impact on prices and firm entry, but also for the commuting and transport costs borne by both firm owners and consumers as a result of the spatial patterns they induce.

Finally, the paper speaks to the development literature on demand constraints to firm growth in low-income countries (Bloom et al., 2010; Hsieh and Olken, 2014). While early research emphasized supply-side barriers, recent work has highlighted the importance of access to larger and more quality-sensitive markets for firm upgrading and expansion (Verhoogen, 2008; Lagakos, 2016; Atkin et al., 2017; Hjort et al., 2020; Goldberg and Reed, 2023; Carrillo et al., 2023). I contribute in two ways. First, I show that firms—particularly higher-quality ones—are willing to incur large fixed costs in the form of commuting to locate in dense consumer markets. Second, I show that consumer search costs suppress overall demand, limiting firm profitability and potentially stunting growth.

The paper is structured as follows. Section 2 introduces the setting of the study and the data. In Section 3, I document key facts about firms' location choices and consumers' search behavior. I rely on this empirical evidence to motivate the structure of the model, which is presented in Section 4. Section 5 describes the estimation procedure, discusses the identification of the model parameters, and summarizes the results of the estimation. These are used to construct counterfactual scenarios in Section 6. Section 7 concludes.

et al. (2023), who show that when consumers must travel to acquire information, spatial concentration can reduce price elasticity and raise mark-ups. The main innovation of this paper is to endogenize firms' location choices in response to these forces.

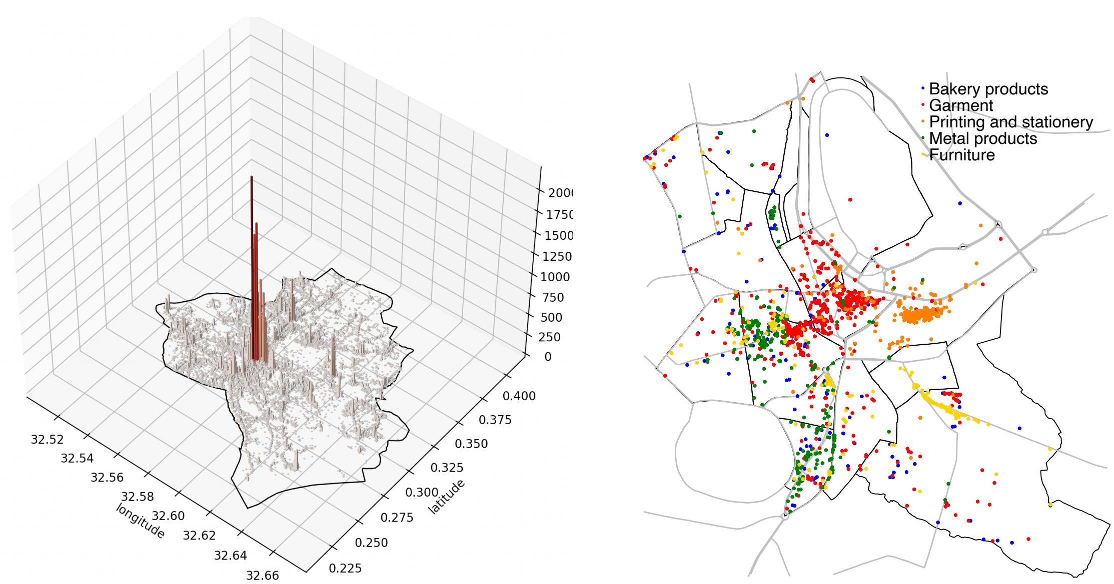
2 Setting and Data

2.1 Kampala Garment Sector

The setting of this study is the garment sector in Kampala. Kampala is the administrative capital and economic hub of Uganda, hosting 29% of all businesses and contributing to 60% of the country's GDP (KCCA, 2019). Panel A of Figure 1 shows the spatial distribution of all formal and informal firms operating in the city. It reveals that economic activity is heavily concentrated on a small area at the heart of city, with 40% of all establishments operating within 2 km from the central business district.[4] Panel B plots the location of all the establishments in this central area for the top five Ugandan manufacturing sectors. Each dot on the map represents a firm and each color corresponds to a four digit ISIC sector. The color pattern indicates that, within this central location, firms cluster by sector, with different areas hosting different industries.

Figure 1: Spatial distribution of firms in Kampala

PANEL A: All firms **PANEL B: Top 5 manufacturing sectors**



Notes: Data is from 2010 Ugandan Census of Business Establishments (UBOS), which covers the universe of formal and informal firms in Uganda. Panel A shows the distribution of all firms in Kampala, with the height of each bar indicating the number of firms within a specific area. Panel B zooms in on the central part of the city and shows the location of all firms in the five manufacturing sectors with the highest number of establishments (and the largest labor force) in Uganda. On the map, each dot is a firm and each color a four digit ISIC sector.

This study focuses specifically on the garment sector. Garment is one of Uganda's key manufacturing industries, accounting for 42% of all manufacturing firms and employing

[4] For comparison, respectively 25% and 11% of firms in London and Los Angeles operate within 2 km of the central business district (author's calculations using CDRC 2021, County Business Patterns 2019).

15% of the manufacturing labor force. Despite its size, the sector is highly fragmented: 77% of businesses consist of a single, self-employed individual, and 84% have an annual turnover below \$2,000.[5] The sector was chosen for this study for two reasons. First, garment firms exhibits strong spatial clustering (Panel B of Figure 1), providing an ideal setting for studying agglomeration forces. Second, garment firms produce goods that are horizontally differentiated. With limited access to information technology, this feature makes in-person search particularly relevant, as it is the only way for customers to discover which firms sell the products that they are looking for.

2.2 Data

Firm sampling The data for this study comes from an original survey of 600 firms and their customers. Firms were selected from an initial listing of all garment businesses operating in one of 14 randomly selected parishes in Kampala. Parish selection was stratified by firm density to include areas with the highest concentration of firms and ensure some variation in density across parishes outside the central part of the city.[6] Figure A2 shows the location of the selected parishes next to a map of the density of garment firms across Kampala. Although the study only covered 14 out of the 96 parishes in the city, these parishes accounted for 68% of all garment firms operating in Kampala according to 2010 Census data.

Interviewers conducted a door-to-door, in-person listing of all the garment firms in the selected parishes, enumerating a total of 2,407 firms. Figure A3 shows the density of firms in each parish based on the listing data. Consistent with previous census data, three parishes in the city center exhibit a significantly higher firm density compared to all other parishes in the sample. For the remainder of the paper, these parishes are referred to as the *core* of the city, while the remaining parishes are termed the *periphery*. From the initial listing, 330 firms in the core and 330 firms in the periphery were selected for the firm survey. Compliance was high, at 89%, and not statistically different between core and periphery, resulting in a final sample of 601 firms. All results in the paper are weighted to reflect the sampling strategy.

[5] These characteristics are by no mean specific to the garment sector. Excluding garment, the median manufacturing firm in Uganda has no employees and has an annual turnover below \$2,000.

[6] The parish is the second lowest administrative unit in Uganda, with an average size of 2.03 square-km. Firm density was measured as the average number of firms per square kilometer in each parish according to 2010 Ugandan Census of Business Establishments. The selection process was as follows. First, parishes with fewer than ten garment firms were excluded from the sample. Second, the remaining parishes were divided into four strata based on firm density: (i) 0-49, (ii) 50-99, (iii) 100-300, and (iv) more than 300 firms per square kilometer. Finally, 4 parishes were randomly selected from stratum (i), 5 from stratum (ii), 3 from stratum (iii) and 2 from stratum (iv) to participate in the study. In the two densest parishes, Nakasero IV and Kisenyi II, interviewers enumerated every second firms they encountered in their random walk.

Customer sampling The list of potential customers was compiled using two data sources. First, interviewers listed all customers who purchased products from the firm during the interview. Second, firms were asked to record their transactions for the three days after the survey. For each transaction, firms recorded the name and contact details of the customer, along with information on whether the customer was a *final consumer*—an individual making a purchase for personal or household use, or a *retailer*—an individual purchasing for resale. In total, the details of 1,510 customers (64% final consumers and 36% retailers) were collected from 385 firms.[7] From this list, 581 customers were randomly selected to take part in the customer survey. The selection was stratified by firm location (core vs. periphery) and type of customer (final vs. retail) to ensure an adequate representation of both customer types across the two locations.

Survey design The data collection was designed with two objectives in mind: (i) understanding how consumers search for products and (ii) exploring the factors influencing firms' location choices. Three data sources contributed to the first objective: transaction data, a customer survey, and a mystery shopper exercise. The firm survey was designed to address the second objective.

Transaction records provided information on the outcome of the search process. These records were obtained by asking owners to maintain a written log of all the transactions conducted by their firm for the three days following the survey.[8] A total of 2,848 transactions were recorded, with information on the type of product (e.g., dress, t-shirt, trousers, skirt), the quantity and price at which it was sold, the type of customer making the purchase—final consumer or retailer—and the location where the customer traveled from. Transaction records were complemented by a customer survey designed to provide a comprehensive view of how customers search for products. This survey included detailed questions on how consumers decide where to look for a firm and which business to buy from once in a location.

The mystery shoppers exercise was aimed at collecting accurate and comparable information on firms' prices and product quality. It involved commissioning the same garment—a dress, the most commonly produced item by firms in the sample—from all firms (Figure A12). An expert tailor designed the dress with specific characteristics to test the tailor's skills.

[7] 64% of firms, equally distributed across core and periphery, provided the contact information of at least one customer. Column 2 of Table A1 shows that attrition was uncorrelated with firm location, number of employees, and monthly revenues, but was positively correlated with the total number of weekly customers reported at baseline.

[8] Firm owners were offered a monetary incentive to keep transaction records. 91% of firms complied. Column 1 of Table A1 shows that attrition was uncorrelated with firm location, number of employees, monthly revenues, and total number of weekly customers. The correlation between the self-reported weekly number of customers in the baseline survey and the transaction records is 0.818.

Interviewers posed as customers and were trained to follow a script (Appendix B.4) to order the dress and negotiate its price. Firms were provided with fabric, an accurate description of the dress, and, upon request, a photo of the product. The expert tailor then anonymously rated the quality of the finished garment according to detailed evaluation criteria listed in Figure A13.

Finally, the firm survey focused on the second objective of the data collection: understanding the drivers of firms' location choices. In addition to standard firm-level information, such as the number of employees, revenues, profits, and characteristics of the firm owner, the survey included questions about the firm's location history and the motivations behind the owner's initial choice of location and any subsequent relocations. Detailed data was also gathered on the firm's production process, with the aim of exploring the potential sources of production externalities.

3 Descriptive Evidence on Firm Location and Search

In this section, I present evidence on firms' location choices and consumers' search behavior in Kampala. I begin by summarizing the characteristics of firms operating in different parts of the city. I then document three facts about the relationship between firm density and consumer search, consistent with the presence of information frictions that are mitigated by higher firm density.

3.1 Firm-Level Differences Between Core and Periphery

Firm characteristics Table 1 presents summary statistics on firm characteristics, disaggregated by location (core vs. periphery). Reflecting broader trends in Uganda's garment sector, firms in my sample are small: the average business employs 1.3 workers (including the owner), owns 2.1 machines—typically a sewing machine and a flat iron—and operates in a space of just 3 square meters. Despite their size, these firms are not recent market entrants: they have been operating for an average of 8 years and report monthly revenues of $167, more than twice the Ugandan monthly GDP per capita of $80. Because of their small scale, these businesses typically function as hybrids between manufacturers and retailers, with the same individual responsible for both production and sales at the same location. As a result, both supply-side and demand-side factors are likely to influence location decisions.

Firms in the core differ significantly from those in the periphery. On average, core firms report monthly revenues and profits that are 78% and 76% higher, respectively, than firms in the periphery. However, they operate with 27% fewer workers, 13% fewer machines, and in spaces half the size. Two factors appear to discourage locating in the core: commuting costs and rents—both approximately double those faced by firms in the periphery.

Table 1: Summary statistics

	All Firms	Core	Periphery	P-value
Number of workers	1.319	1.250	1.701	[0.000]
	(1.239)	(1.172)	(1.505)	
Number of machines	2.142	2.092	2.418	[0.075]
	(2.150)	(2.155)	(2.114)	
Size of premises (square meter)	3.005	2.652	4.952	[0.000]
	(4.144)	(3.568)	(6.117)	
Firm age (years)	8.001	7.974	8.151	[0.814]
	(8.019)	(7.701)	(9.611)	
Monthly revenues (USD)	167.039	179.402	100.611	[0.000]
	(172.362)	(180.143)	(100.401)	
Monthly profits (USD)	44.693	48.002	27.289	[0.000]
	(44.340)	(43.806)	(43.312)	
Rent per square meter (USD)	19.459	20.847	11.717	[0.000]
	(14.211)	(13.999)	(12.910)	
Monthly commuting cost (USD)	36.642	39.817	19.564	[0.000]
	(21.122)	(18.935)	(23.979)	
Number of observations	601	302	299	

Note: Data is from the firm survey. The table reports means, standard deviations (in parenthesis), and p-values from a t-test of equality of means between the core and the periphery (in brackets). All estimates are weighted to be representative of the universe of garment firms in the sampled parishes.

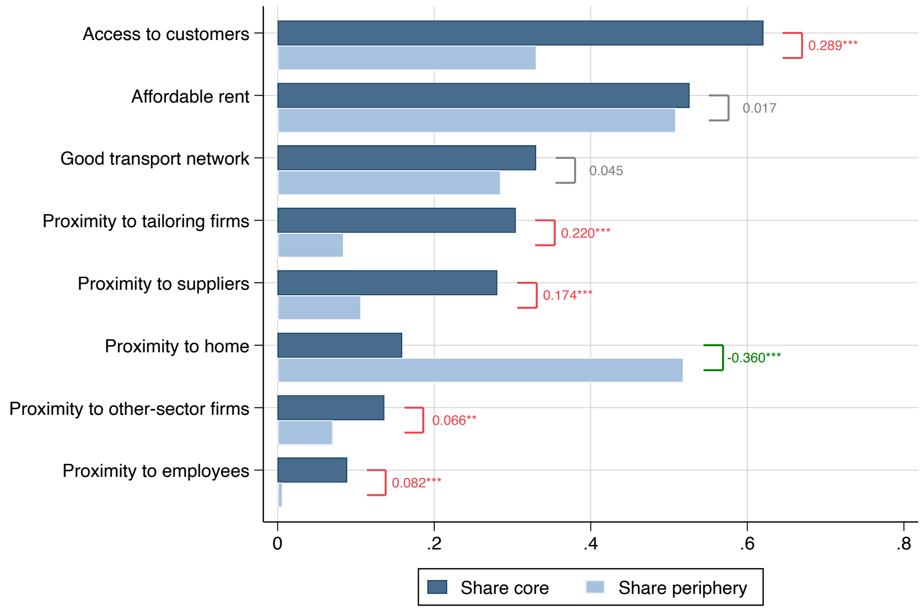
Determinants of location decision What drives firms' initial location decisions? To explore this, I asked firm owners (i) what constraints they faced when starting their business and (ii) what factors shaped their location choice.[9] Responses to both questions consistently highlight demand considerations. When asked about startup constraints, the most frequently cited was "finding customers," mentioned by 73% of firms. In comparison, access to finance, a widely studied constraint in low-income settings, was cited by only 53%. Transport costs were the third most common constraint (11.4%), suggesting that commuting distances also influences location choices.[10]

[9] Location decisions are highly persistent: Table A2 shows that 54% of firms have never moved, and 23% have only moved within the same area (core or periphery). Just 5% and 3% report relocating from the periphery to the core and vice versa, respectively. The remaining 14% moved from outside Kampala.

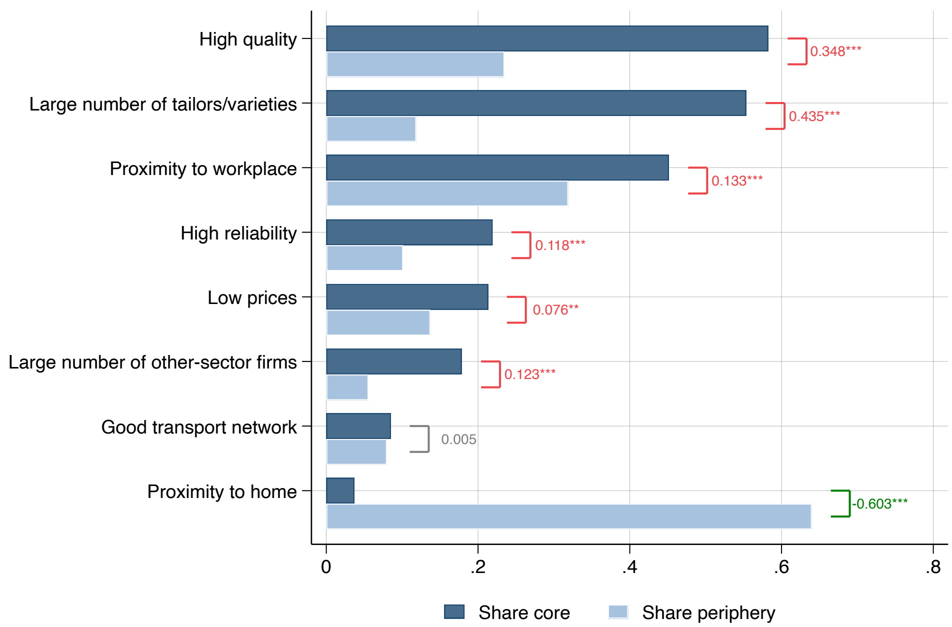
[10] Other cited startup constraints include high taxes/license fees (10%), difficulty finding suppliers (9.9%), lack of managerial skills (8.1%), intense competition (7.1%), limited space (6.1%), difficulty accessing machines (4.9%), and high rental costs (4.8%).

Figure 2: Reasons for choosing core vs. periphery

Panel A: Firms



Panel B: Customers



Notes: Data is from firm and the customer surveys. In Panel A, the dark blue bars show the share of firms in the core reporting the reason indicated on the left as a driver of their initial choice of location. The light blue bars show the same statistic, but for firms operating in the periphery. In Panel B, the dark blue bars show, among customers who indicated they would prefer to search for a new firm in the core, the share reporting the reason indicated on the left as a driver of their choice of location. The light blue bars show the same statistic, but for customers who indicated they would prefer to search for a firm in the periphery. In both panels, the numbers next to the bars are the differences between the share of firms/customers in the core and the periphery reporting a given reason, with the stars indicating their level of significance.

Figure 2, Panel A, further illustrates the drivers of firms' location choices. Specifically, it shows: in dark blue, the share of firms in the core citing a particular reason as a determinant of their location decision, and in light blue, the corresponding share for firms in the periphery. The numbers on the figure indicate the difference between the shares of core and periphery firms reporting each reason, along with the corresponding significance level.

Two key insights emerge from this graph. First, the primary reason firms choose to locate in the core is to access customers: 62% cite this as a key factor in their location decision, pointing to the large number of individuals shopping in the area. Second, while standard agglomeration benefits—such as proximity to suppliers (28%) and access to labor (9%)—do matter, they appear to be less central. In contrast, the main pull factor for locating in the periphery is proximity to home (52% vs. 16% in the core), highlighting commuting costs as a key congestion force. Additional potential reasons, such as price collusion, bargaining, proximity to other sector firms, and gender preferences for commuting are discussed in Appendix B.5.

Table 2: Transaction characteristics in core and periphery

	(1) Daily revenues (USD)	(2) Number of daily customers	(3) Share of retail customers	(4) Transaction Value (USD)	(5) Quantity	(6) Unit price (USD)
Panel A: No Controls						
Core	9.336***	-0.163**	0.446***	8.497***	12.23***	1.848***
	(2.340)	(0.080)	(0.029)	(0.932)	(1.125)	(0.326)
Panel B: Product FEs						
Core				4.277***	14.80***	-0.079
				(0.860)	(1.422)	(0.243)
Panel C: Product FEs and Quality						
Core				4.173***	14.51***	-0.258
				(0.920)	(1.494)	(0.245)
Quality score				0.566	-1.043	0.511**
				(0.611)	(0.928)	(0.241)
Mean in periphery	7.423	0.980	0.102	6.763	3.628	3.136
Number of observations	546	546	512	2,726	2,726	2,458

Note: * $p < .10$, ** $p < .05$, *** $p < .01$. Data is from transactions records and the mystery shoppers exercise (for quality scores). In Columns 1 to 3, the unit of observation is the firm. In Columns 4 to 6, it is the transaction. Panel A shows the results from a regression of the outcomes on a dummy equal to one for firms in the core, without any additional control. Panel B adds product type fixed effects. Panel C controls for quality scores in addition to product fixed effects. In Column 5, all regressions control for quantity purchased in the transaction. The mean value of the outcome in the periphery is shown at the foot of the table.

Transaction characteristics I now turn to transaction-level data to understand the sources of higher revenues in the core. Table 2 reports results from regressions of various transaction characteristics on a dummy for being located in the core. It shows that firms in the core generate significantly higher revenues than those in the periphery, despite serving 18% fewer daily customers. The revenue difference is largely explained by differences in customer type: core firms serve a much higher share of retailers (55% vs. 10% in the periphery). This matters because retailers make larger purchases. As Column 5 shows, the median transaction in the periphery—typically made by a final consumer—involves one unit, while in the core, the median (retailer) transaction involves ten units (mean: 3.6 vs. 16 units). As a result, for the same product type, the average transaction value is 65% higher in the core (Column 4, Panel B).

This quantity difference is not explained by variation in product types, quality, or prices. The coefficient on the core dummy in Column 5 remains largely unchanged after controlling for product fixed effects (Panel B) and the quality score from the mystery shopper data (Panel C). In addition, Column 6 shows no significant price difference between locations after controlling for product type, quality, and quantity purchased (Panel C).[11] Instead, the data suggests that bulk buyers benefit from economies of scale in transport. I provide additional evidence for this in the next section, where I analyze consumer search behavior.

3.2 Stylized Facts about Search

Having discussed the determinants of firms' location choices, I now turn to the behavior of customers. This section presents three facts about the relationship between firm density and consumer search that guide the structure of the quantitative model.

FACT 1: Consumer search is mainly in person and entails substantial costs

Search in this setting overwhelmingly takes place offline. 91% of all transactions occur in person, with walk-ins being the most common method through which customers identify new firms (55%)[12] and internet being the least common (6%). Customers spend the equivalent of 8.6% of the transaction value on transport to the firm. Travel to the core is particularly costly: transport expenses are nearly three times higher than to the periphery ($1.25 vs. $0.48), excluding the opportunity cost of time, which averages 33 minutes for a one-way trip to the core and 17 minutes to the periphery. Two factors explain these higher costs. First,

[11] Mystery shopper data and self-reported data on the price of the most common product sold by the firm also shows no significant price differences across locations (p-values = 0.347 and 0.276 respectively). This is consistent with the fact that customers rarely cite prices as a reason for choosing where to shop.

[12] Table A3 shows that among final consumers, walk-ins are followed by asking friends or family for recommendations (43%). Retailers are instead more likely to rely on recommendations from other firms (51%). Interestingly, the number of firms visited by customers who receive a recommendation prior to purchasing is not statistically different from those visited by individuals who search randomly, suggesting that even with referrals, customers still engage in independent search.

while 59.2% of firms are concentrated in the core, the population is more evenly distributed across the city, with less than 1% residing in the core (Figure A7). Second, when shopping in the periphery, customers typically buy products from firms located within their own parish or a nearby area (Figure A8).

These high costs are not borne equally by different types of customers. Importantly, the data shows that there is a negative relationship between quantity purchased and unit transport cost to the firm location (Panel A of Figure A9). For the lowest quintile of the purchased quantity distribution, average transport costs to the firm amount to 32% of the transaction value. For the highest quintile, these costs only correspond to 5.5% of the transaction value. This pattern holds despite the fact that customers buying larger quantities travel further to source their products (Panel B of Figure A9), suggesting that transport costs are fixed and generate to economies of scale in transport (Grant and Startz, 2022).

FACT 2: In denser locations, consumers search more and more easily find the products they seek

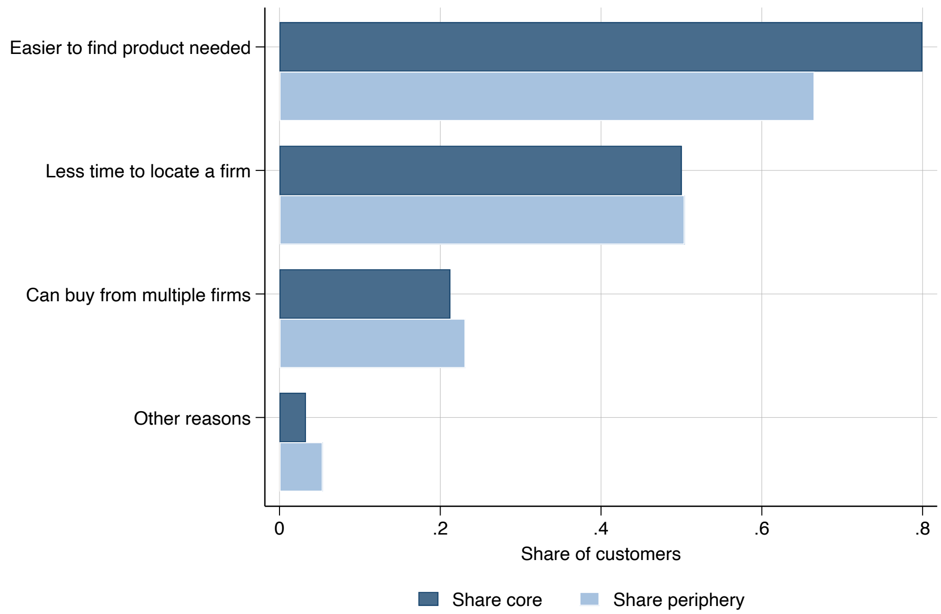
Despite the high transport costs documented earlier, many consumers still travel to the dense core. Survey responses reveal why. Among those searching in the periphery, 64% report doing so to stay close to home and save on transport. By contrast, customers searching in the core emphasize three distinct advantages: 58% cite firms' reputation for quality, 55% the availability of many tailors and product varieties, and 45% proximity to the workplace (Figure 2, Panel B).[13]

In relative terms, access to a large number of tailors is the most important factor explaining why customers prefer the core to the periphery. When asked why they prefer areas with many tailoring firms, 80% of individuals searching in the core reported that density makes it easier to find the product they need, while 50% said it reduces the time it takes to locate a firm (Figure 3). Both responses highlight density as a mechanism that lowers search costs for consumers with limited information about product types and firms' exact location. Density

[13] Only 18% of customers cite the presence of other-sector firms as a reason for searching in the core, and 83.2% report purchasing from just one garment firm when visiting this location. This suggests that trip chaining—the tendency to make purchases from different firms in the same trip (Miyauchi et al., 2021; Oh and Seo, 2022)—plays a limited role relative to information frictions. Nevertheless, this channel is incorporated in the structural model. Low prices, mentioned by 21% and 14% of customers in the core and the periphery, also appear to have little influence on customers' preference for the core. This aligns with evidence from the mystery shoppers exercise (Appendix B.5) showing that (i) there are no significant price differences across locations and (ii) customers are not able to bargain for better prices in the core. Figure A6 replicates Panel B of Figure 2 separately for final consumers and retailers. The main difference is that, in absolute terms, proximity to the workplace matters more for retailers, while proximity to home matters more for final consumers. In the model, I assume that the origin location of final consumers is their home, while for retailers it is their business.

also eases search in the periphery, but, as shown in Figure 2, far fewer customers cite it as a reason for searching in that area.[14]

Figure 3: Reasons for preferring locations with a large number of firms



Notes: Data is from the customer survey. Figure 3 shows share of customers reporting a given reason for preferring to search areas with a large number of tailoring firms. The dark blue bars show the share of customer searching in the core, while the light blue bars the share of customers searching in the periphery citing a specific reason.

Observed behavior reinforces this interpretation. First, the data show that, when searching for a new supplier, customers visit multiple firms prior to purchasing (mean = 1.95). This is inconsistent with perfect information, which would eliminate the need for search. Second, customers searching in the core visit, on average, 31% more businesses before purchasing (2.39 vs. 1.82; p-value = 0.004; Table A4), a result that is robust to controlling for quality scores, prices, customer type (final consumer vs. retailer), and product type fixed effects. These figures almost certainly understate the extent of search in dense areas, where many firms operate side by side in the same room, making it difficult for customers to distinguish between them. Indeed, firm-level GPS data show that the average business in the core is surrounded by 8.7 other firms within a 10-meter radius, compared to just 0.5 in the periphery.

Taken together, this evidence highlights two central facts. First, consumers engage in costly search by visiting multiple firms before buying, which would not occur if they had full information. Second, firm density lowers these costs by making it cheaper to observe and compare a larger number of products in one trip. This suggests that when searching in the

[14] In line with the fact that 83.2% of consumers report purchasing from only one garment firm per trip, trip chaining emerges as a secondary benefit of density, with just 22% mentioning it as a reason for preferring dense areas.

core, customers trade off higher transport costs *to* the location for lower search costs *within* the location.

FACT 3: In the core, customers find higher quality products and better matches

Table 3 reports correlations between an indicator for firms being in the core and four dimensions of quality: (i) tailoring quality, summarized by the expert tailor's score; (ii) material quality, reflected in the cost of the fabric used in production; (iii) supplier reliability, measured by whether the firm delivered the product on time in the mystery shoppers exercise; and (iv) match quality, proxied by the average length of the firm-customer relationship.

Table 3: Differences in quality between core and periphery

	(1) Quality score	(2) Price of 1 m of fabric	(3) Delivered on time	(4) Length of relationship
Core	0.198**	0.541***	-0.037	0.993**
	(0.100)	(0.189)	(0.038)	(0.458)
Quality score				-0.108
				(0.210)
Price of 1 m of fabric				-0.025
				(0.106)
Delivered on time				0.997*
				(0.550)
Interviewer FEs	✓	✓	✓	
Product and Customer FEs				✓
Mean in periphery	-0.078	2.871	0.807	2.137
Number of observations	494	476	494	494

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. Robust standard errors in parentheses. The table shows the results from a regression of the three quality measures on a dummy equal to one for firms in the core. The Quality Score score is a standardized measure with mean 0 and variance 1. The length of the relationship is measured in years. The regression in Column 4 includes the age of the firm as a control variable.

Columns 1 to 3 show that the quality of goods sold in the core is higher than in the periphery, both in terms of tailoring and materials used. There are two possible explanations for this: either higher-quality firms sort into the core, or learning is more likely to occur in denser areas. To test the latter hypothesis, I repeat the analysis controlling for the owner's years of experience and an interaction between experience and being located in the core. The results in Table A5 show no significant difference in the experience gradient across locations. This is inconsistent with greater learning in the core and suggests instead that quality differences are driven by selection, a mechanism incorporated in the model.

Column 4 examines the duration of firm-customer relationships across locations. Two points stand out. First, the average relationship length in this setting is two and a half years, consistent with customers engaging in costly search to find long-term suppliers. Second, relationships in the core last, on average, one year (46%) longer than those in the periphery. This result is robust to controlling for the three measures of quality in Columns 1 to 3, firm age, and customer and product type fixed effects. The finding highlights that the firm-customer match, rather than just the vertical quality of the product, is an important driver of search. It also connects directly to the previous fact: density reduces within-location search costs, enabling customers to compare products across more firms, find the variety needed and ultimately form longer, more durable relationships.

4 Model

In this section, I present a model of consumer search and firm location in the presence of information frictions. The key mechanisms are guided by the evidence presented in Section 3. Uninformed consumers choose where to search for goods, weighting transport costs *to* a location against search costs *within* it. In locations with a high density of firms, consumers face lower search costs, which enables them to observe more product varieties and find better matches. High-density areas, typically further from consumers, particularly attract customers buying in bulk, who benefit from economies of scale in transport.

I develop the quantitative framework in three steps. First, I build a discrete choice model of demand that incorporates imperfect information and economies of scale in transport. I show that, by generating consumption externalities, these mechanisms can lead to greater demand for firms in denser areas. Second, I model firms' pricing and input choices explicitly, allowing for productivity to depend on local firm density in order to capture standard production externalities. Finally, I analyze how consumption and production externalities interact to shape firms' location decisions, which I model as a static entry game.

4.1 Setup

The economy consists of L locations (indexed $l = 1, 2, ..., L$) and features a discrete number of firms ($j = 1, 2, ..., I$) and consumers ($i = 1, 2, ..., J$).

Firms are single-product,[15] produce differentiated goods, and have idiosyncratic preferences over locations. They decide: (i) where to locate; (ii) what price to charge; (iii) what combination of land and labor to employ.

Consumers purchase one type of good,[16] but are heterogeneous in the quantity demanded

[15] Transaction data show that on average 75% of a firm's transactions involve the sale of one product.

[16] Transaction data shows that 92% of customers only purchase one type of good in a given transaction.

of that good. They have idiosyncratic preferences over products and idiosyncratic search costs both across and within locations. They choose: (i) in which location to search; (ii) how many firms to visit within that location; and (iii) from which firm to purchase.

The model is static. This simplifying assumption is supported by the data, which show that both firm and consumer choices are highly persistent, and it also keeps the computation of a spatial equilibrium with many firms and locations tractable.[17] The framework should therefore be viewed as modeling the formation of durable firm-customer relationships: customers search once to form a match with a firm, and this match persists indefinitely. Firms, in turn, choose their location in anticipation of the long-run value of the matches they will be able to form in each location.

4.2 Demand

Utility The indirect utility of consumer i purchasing q_i units of product j in location l is:

$$u_{ijl} = q_i^{\theta}\Big(\beta \mathbf{x}_j + \xi_j + (1-\sigma)\varepsilon_{ij}\Big) - \alpha\, p_{jl} q_i + \gamma_l \tag{1}$$

where $\mathbf{x}_j$ and ξ_j are product j's observable and unobservable (to the econometrician) quality attributes, ε_{ij} is an idiosyncratic match value, p_{jl} is the price, q_i is the consumer's quantity demanded, and γ_l is a fixed effect for locations with a high concentration of firms across all sectors.[18]

All consumers have the same ranking over $\mathbf{x}_j$ and ξ_j, which represent the *vertical differentiation* across products. By contrast, ε_{ij} is a match parameter that captures the fit between consumer i's preferences (needs) and the specific variety offered by firm j (*horizontal differentiation*). I assume that ε_{ij} is only observed upon visiting a firm and is distributed as a standard Type I Extreme Value, with $\sigma \in [0, 1]$ governing its variability within a location. Quantity demanded q_i is heterogeneous across consumers and treated as an exogenous consumer type.

For ease of exposition, I assume a discrete number of firm types N_j, indexed by j. Firms of the same type share identical observable and unobservable quality $\mathbf{x}_j$ and ξ_j, but offer differentiated varieties through distinct draws of ε_{ij}. Let N_{jl} denote the number of type-j firms operating in location l, and $N_l = \sum_{j=1}^{N_j} N_{jl}$ the total number of firms in location l.

Search When choosing where to search, consumers know the distributions of $\mathbf{x}_j$ and ξ_j and have correct conjectures about equilibrium prices (Wolinsky, 1983; Moraga-González et al.,

[17] In Appendix C.2, I extend the model to include a second period and show that the results are qualitatively similar.

[18] This term may capture consumers' preferences for trip chaining across sectors or proximity to their workplaces.

2023). Once in a location, they observe $\mathbf{x}_j$ and ξ_j for all firms,[19] but must visit the firm in person to learn the value of ε_{ij}.

The timing of the consumer decision is as follows. First, consumers choose in which location to search based on the expected utility, the transport costs and the search costs associated with each location.[20] Second, once in a location, consumers decide in what order to visit firms and pay a within-location, firm-specific search cost for each visit. After visiting a firm, they decide whether to purchase one of the products observed so far or continue searching. Consumers also have the outside option of not searching. However, conditional on traveling to a location, they must buy one of the products available there.[21] I normalize the utility of the outside option u_0 to be zero for final customers, but allow retailers to have a different outside option, which I estimate.

The cost of consumer i visiting location l takes the following form:

$$C_{il} = \tau_1 g_1(||z_i - z_l||) - \omega_{il} \tag{2}$$

where $\tau_1 g_1(||z_i - z_l||)$ is the transport cost of traveling to location l, and ω_{il} is a *consumer-location* specific search cost that may capture idiosyncratic information an individual has about a particular area. Transport costs depend on the distance between the consumer and the location, while ω_{il} is iid and distributed as a standard Type I Extreme Value. Importantly, C_{il} does not depend the quantity purchased, which embeds the idea of this type of costs being fixed.

Once in a location, consumers pay an idiosyncratic within-location search cost, c_{ij}, to visit each firm. This cost is *consumer-firm* specific and may reflect, for instance, recommendations from other individuals or heterogeneity in the exact point where the consumer begins searching within the area. I assume c_{ij} follows a cumulative distribution function F^c_{ijl}, which depends on firm density $\frac{N_l}{a_l}$. In a setting where search occurs in person, this captures the idea that searching is less costly in locations where firms are close to one another.

[19] This assumption is consistent with Table A4, which shows that the number of firms visited prior to purchase is uncorrelated with quality and prices, suggesting that these are not the product dimensions customers discover through search. The assumption also improves the tractability of the problem given the large number of firms in this setting.

[20] I assume that consumers cannot visit more than one location. This is consistent with survey data showing that, among customers with more than one supplier, only 21% purchase from firms in both the core and the periphery.

[21] This assumption is mainly due to the lack of data on customers who visit a location without making a purchase.

4.2.1 Within Location Search

I consider the within-location search process of consumer i, who has chosen to visit location l. This process closely follows the setup of Moraga-González et al. (2023). Within a location, search is sequential, ordered, and with costless recall. Since ε_{ij} is iid across firms and consumers, Weitzman (1979) solution can be applied to characterize optimal consumer search. Specifially, Weitzman (1979) shows that the optimal search strategy for consumer i in this set-up is to (i) visit sellers in decreasing order of reservation value and (ii) stop searching once the highest realized utility exceeds the reservation value of the next-best option to be searched.

Given the assumed distribution of ε_{ij}, consumer i's utility from product j in location l is distributed as:

$$F_{ijl}(z) = \exp\left(- \exp\left(\frac{\delta_{ijl} - z}{q_i^{\theta}(1-\sigma)} \right) \right) \tag{3}$$

where $\delta_{ijl} = q_i^{\theta}(\beta \mathbf{x}_j + \xi_j - \alpha p_{jl} q_i^{1-\theta})$ denotes the mean utility from purchasing q_i units of the good. Let $H_{ijl}(r) = \int_r^{\infty}(z-r)dF_{ijl}(z)$ denote the expected gains from searching an additional firm when the best utility realized so far is r. The reservation value is defined as the utility level that equates expected gains with the marginal cost of searching:

$$\int_r^{\infty} (z - r)\, dF_{ijl}(z) = c_{ij}\left(\tfrac{N_l}{a_l} \right) \tag{4}$$

Since the left hand side of equation (4) is decreasing in r and strictly convex, the equation admits a unique solution. Let $r_{ijl} = H_{ijl}^{-1}(c_{ij}(N_l/a_l))$ denote the reservation value of consumer i associated with searching firm j in location l. It is straightforward to see that if $c'_{ij}(N_l/a_l) < 0$—that is, if search costs decrease with firm density—then reservation values r_{ijl} increase with firm density. This implies that, all else equal, consumers will visit more firms in high-density locations, consistent with *Fact 2* presented in Section 3.

The large number of firms and the heterogeneity across locations, firms, and consumers imply that computing demand for each seller would require tracing an extensive set of purchasing paths, making Weitzman's solution impractical to implement. I therefore rely on a result from Armstrong (2017) and Choi et al. (2018), which translates Weitzman's characterization into a simple discrete-choice framework. Let $w_{ijl} = \min\{r_{ijl}, u_{ijl}\}$. Armstrong (2017) and Choi et al. (2018) show that the optimal solution to the sequential search problem is equivalent to consumers choosing the firm with the highest w_{ijl}. If search costs are equal to zero, this reduces to the full-information choice: consumers purchase the product yielding the highest realized utility. Conversely, as search costs increase, consumers search less and rely more on firms' observed characteristics and idiosyncratic search costs to make their decision.

Following Moraga-González et al. (2023), I propose a parametric distribution of search costs that yields closed-form expressions for product purchase probabilities. This distribution, which is derived in Appendix A.1, has the following CDF:

$$F^c_{ijl}(c) = \frac{1 - \exp\left(-\exp\left(\frac{-\left(H_0^{-1}(c)+\mu\left(\frac{N_l}{a_l}\right)\right)}{q_i^\theta(1-\sigma)}\right)\right)}{1 - \exp\left(-\exp\left(\frac{-H_0^{-1}(c)}{q_i^\theta(1-\sigma)}\right)\right)} \tag{5}$$

Given this functional form assumption for the search cost distribution, I show in Appendix A.1 that w_{ijl} is distributed as a Type I Extreme Value with location parameter $\delta_{ijl} - \mu(\frac{N_l}{a_l})$ and scale parameter $q_i^\theta(1-\sigma)$.[22] The probability that consumer i purchases product j, conditional on visiting location l, is therefore:

$$P_{ij|l} = \Pr\left(w_{ijl} > w_{ij'l} \forall j' \neq j\right) = \frac{\exp\left(\frac{\delta_{ijl}-\mu(\frac{N_l}{a_l})}{q_i^\theta(1-\sigma)}\right)}{\sum_{j'=1}^{N_j} N_{j'l}\exp\left(\frac{\delta_{ij'l}-\mu(\frac{N_l}{a_l})}{q_i^\theta(1-\sigma)}\right)} \tag{6}$$

This expression captures firm competition *within* a location. All else equal, the share of customers purchasing from firm j, conditional on searching in location l, decreases with the number of type-j firms operating in that location (N_{jl}). This result is intuitive: holding the pool of customers visiting location l fixed, an additional firm expands the choice set, reducing each firm's share. I refer to this business-stealing effect as the *market-share effect*.

4.2.2 Across Location Search

Consumers choose the location that maximizes their expected utility V_{il}. Given the distribution of w_{ijl}, consumer i's expected utility from location l is:

$$V_{il} = E_w\left[\max_{j\in l}\left(w_{ijl} - C_{il}\right)\right] = q_i^\theta(1-\sigma)\log\left(\sum_{j=1}^{N_j} N_{jl}\exp\left(\frac{\delta_{ijl} - \mu\left(\frac{N_l}{a_l}\right)}{q_i^\theta(1-\sigma)}\right)\right) + \gamma_l - C_{il} + \kappa_i \tag{7}$$

where C_{il} is the search cost defined in equation (2) and $\kappa_i = q_i^\theta(1-\sigma)\gamma$ is a consumer type-specific constant, with γ the Euler-Mascheroni constant. The main takeaway from this expression is that, all else equal, expected utility from a location increases with the number of type-j firms in that location (N_{jl}). Two mechanisms drive this effect. The first one is a

[22]For F^c_{ijl} to be a proper distribution, $\mu(\frac{N_l}{a_l})$ must be positive. I therefore assume that it takes the following functional form: $\mu(\frac{N_l}{a_l}) = \log(1+\exp(\mu\frac{N_l}{a_l}))$. Under this assumption, one can verify that the search cost distribution in equation (5) is increasing in c and takes value 1 as c approaches infinity.

shift in the search cost distribution, captured by the parameter $\mu(\frac{N_l}{a_l})$. If average search costs decrease with density, i.e. $\mu'(\frac{N_l}{a_l}) < 0$, consumers optimally visit more firms and observe more product varieties in denser locations. In expectation, this raises the likelihood of finding a good match (a high ε_{ij} draw), consistent with *Fact 3* presented in the previous section.

The second channel is an increase in the number of search cost draws, reflected in the N_{jl} term in the summation. In the model, idiosyncratic within-location search costs c_{ijl} are observed only upon visiting the location. A larger number of firms increases the number of draws, lowering the expected minimum value of c_{ijl}. This parallels the love of variety property in logit-type discrete choice models with perfect information, where expected utility rises with the number of available varieties (Anderson et al., 1992). In this model, however, access to varieties is mediated by firm-specific search costs.

Given the assumed distribution of ω_{il}, the probability of consumer i purchasing in location l is:

$$P_{il} = \frac{\left(\sum_{j=1}^{N_j} N_{jl} \exp\left(\frac{\delta_{ijl}}{q_i^{\theta}(1-\sigma)}\right) \right)^{q_i^{\theta}(1-\sigma)} \exp\left(-\mu\left(\frac{N_l}{a_l}\right) - \tau_1 g_1(\|z_i - z_l\|) + \gamma_l \right)}{\exp(u_{0i}) + \sum_{l=1}^{L} \left(\sum_{j=1}^{N_j} N_{jl} \exp\left(\frac{\delta_{ijl}}{q_i^{\theta}(1-\sigma)}\right) \right)^{q_i^{\theta}(1-\sigma)} \exp\left(-\mu\left(\frac{N_l}{a_l}\right) - \tau_1 g_1(\|z_i - z_l\|) + \gamma_l \right)} \tag{8}$$

This expression highlights the forces affecting firm competition *across* locations. First, the probability of consumers visiting location l rises with product quality ($\mathbf{x}_j$ and ξ_j) and falls with prices (p_{ijl}) and travel distance ($||z_i - z_l||$). Second, both the summation and the search cost parameter in the numerator capture the agglomeration forces arising from within-location search. By offering lower average search costs, locations with a high N_{jl} attract a greater share of customers. I refer to this mechanism as the *market-size effect*.

4.2.3 Unconditional Demand

Expressions (6) and (8) show that, at a given set of prices, the presence of an additional firm affects demand through two channels: (i) it attracts customers to the location by lowering search costs (*market-size effect*); and (ii) it intensifies competition within the location (*market-share effect*). The unconditional probability of consumer i purchasing product j in location l is the product of these two expressions: $P_{ijl} = P_{il} \times P_{ij|l}$.[23] Hence, the overall impact of N_{jl} on demand depends on the relative strength of the market-size and the

[23] Aggregating over all consumer types, the overall demand for firm j in location l is given by: $Q_{jl}(\mathbf{L},\mathbf{p}) = \int q_i P_{ijl}(\mathbf{L},\mathbf{p}) dG(q, z)$, where $dG(\cdot)$ is the exogenous joint distribution of customers' types and origin.

market-share effects.

To illustrate this more clearly, I temporarily ignore the integer constraint on N_{jl} and assume that prices are independent of firms' spatial distribution. Under these assumptions, the marginal effect of the number of type-j' firms in location l on the unconditional demand for type-j firms in the same location is:

$$\frac{\partial P_{ijl}}{\partial N_{j'l}} = P_{il}P_{ij|l}\left[\left(P_{ij'|l}q_i^{\theta}(1-\sigma) - \mu'\left(\frac{N_l}{a_l}\right)\right)(1-P_{il}) - P_{ij'|l}\right] \tag{9}$$

As expected, the sign of this expression is ambiguous. In particular, the marginal effect in equation (9) is more likely to be positive: (i) the larger the quantity q_i purchased by consumers, (ii) the higher the dispersion of firm-specific taste shocks within a location (low σ), and (iii) the lower the share of consumers purchasing in the location, P_{il}.

First, consumption externalities are stronger for bulk buyers, since the utility gain from finding a better match is realized over all units purchased. This effect is not offset by higher transport costs to dense locations, as these costs are independent of quantity. As a result, larger buyers are more likely to purchase from high-density locations, consistent with the evidence presented in the Section 3.

Second, a lower σ implies greater product differentiation within a location. From the customer's perspective, this raises the marginal value of observing more products, making locations with a larger number of firms relatively more attractive.

Third, although equation (9) shows that $\frac{\partial P_{ijl}}{\partial N_{jl}}$ is non-monotonic in P_{il}, the effect eventually turns negative as P_{il} approaches one. Intuitively, if a location is already highly attractive—because it hosts a large number of firms, offers high-quality products at low prices, or is geographically close to consumers—then an additional firm increases its relative attractiveness only marginally.

Finally, equation (9) implies that the magnitude of the marginal effect increases with $P_{ij|l}$. This means that when the effect of an additional firm is positive, firms with a higher mean utility δ_{ijl} benefit disproportionately from agglomeration. As a result, in the presence of strong consumption externalities, high-quality firms are more likely to sort into denser locations, consistent with *Fact 3* in the previous section.

4.3 Supply

Production Firms produce output using labor ℓ and land h according to a Cobb–Douglas, constant-returns-to-scale production function: $f(h, \ell) = A_l \ell^{\delta} h^{1-\delta}$. I allow for heterogeneity across locations in productivity A_l, wages w_l, and rents r_l, but assume that firms within a location are equally productive. To capture production externalities arising from proximity to suppliers, thick labor markets, or knowledge spillovers, I model productivity as log-linear in density following Combes and Gobillon (2015): $\log(A_l) = T_1 + T_2 \log(\frac{N_l}{a_l})$.[24] In the data, production externalities are not separately identified from other determinants of productivity differences across locations, such as amenities or unobserved sorting (Combes et al., 2012). In Section 5, I argue that attributing all productivity differences across locations to production externalities provides an upper bound on the strength of this channel.

Pricing Conditional on their location choices, firms set prices simultaneously in a static Bertrand game. Optimal prices are implicitly defined by the first order condition below (see Appendix A.2 for a derivation):

$$p_{jl}^{*} = c_l + \frac{1-\sigma}{\alpha}\left(\frac{\int q_i P_{ijl}, dG(q,z)}{\int q_i^2 P_{ijl}\big(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta})\big), dG(q,z)}\right) \tag{10}$$

The system of best-response functions can be written compactly as: $\mathbf{p} = \mathbf{c} - \Lambda(\mathbf{p})^{-1}Q(\mathbf{p})$, where Λ is the $J \times J$ matrix of price derivatives (Berry, 1994). A Nash equilibrium of this game is a vector $\mathbf{p}^*$ that solves this system of equations. In Appendix A.3, I follow Mizuno (2003) to derive conditions for existence and uniqueness of a price equilibrium.

The effect of the number of firms in a location on prices is ambiguous. As shown in Appendix A.4, the sign of this effect depends on the relative strength of three factors: (i) production externalities, which reduce marginal costs in denser locations, lowering prices; (ii) the market-share effect, which intensifies competition within dense locations, leading firms to lower prices; and (iii) the market-size effect, which increases the attractiveness of larger locations to consumers, enabling firms to raise mark-ups. This last mechanism is analogous to the findings of (Murry and Zhou, 2020) and (Moraga-González et al., 2023), who show that in the presence of search frictions, agglomeration can reduce demand elasticity and increase price-cost margins.

4.4 Location

I model firms' location decisions as a static entry game. Firm owners simultaneously choose where to locate their business and can enter only one location. Let $l \in \{1, 2, ..., L\}$ denote the set of choice alternatives. Owners choose location to maximize the following profit function:

[24] Isolating a single source of production externalities is beyond the scope of this paper. However, in Appendix B.6 I show that one potential source of production externalities is outsourcing. Data from the firm survey indicate that firms in the core are more likely to outsource parts of the production to firms specialized in particular tasks, which are themselves concentrated in the core. These firm-to-firm interactions resemble the machine rental market studied by Bassi et al. (2022b) in the Ugandan carpentry sector.

$$\Pi_{jl}(\mathbf{L},\mathbf{p}) = \pi_{jl}(\mathbf{L},\mathbf{p}) - \tau_2 g_2(||z_j - z_l||) + e_{jl} \tag{11}$$

where $\pi_{jl}(\mathbf{L},\mathbf{p})$ denotes the variable profits of firm j in location l, and $\mathbf{L} = \{N_1, N_2, ..., N_L\}$ is the vector of firm counts across locations. To enter a location, owners incur a commuting cost $\tau_2 g_2(||z_j - z_l||)$, which depends on the distance between the owner's residence (or workplace) and the firm.[25] The term e_{jl} is an idiosyncratic entry or set-up cost, which I assume to be iid across firms and locations, distributed as a Type I Extreme Value with scale parameter λ, and known to all firms. Under this assumption, the probability that firm j chooses location l is given by:

$$\Psi_{jl}(\mathbf{L},\mathbf{p}) = \frac{\exp\left(\frac{\pi_{jl}(\mathbf{L},\mathbf{p}) - \tau_2 g_2(||z_j - z_l||)}{\lambda}\right)}{\sum_{l=1}^{L} \exp\left(\frac{\pi_{jl}(\mathbf{L},\mathbf{p}) - \tau_2 g_2(||z_j - z_l||)}{\lambda}\right)} \tag{12}$$

where the number of firms in location l is $N_l = \sum_j \Psi_{jl} N_j$.

Externalities Consumption and production externalities enter the firm's location choice through their effect on variable profits. Taking the total derivative of profits with respect to the number of firms in location l yields:

$$\frac{d\Pi_{jl}(\mathbf{L},\mathbf{p})}{dN_l} = \underbrace{(p_{jl} - c_l)\frac{\partial Q_{jl}(\mathbf{L},\mathbf{p})}{\partial N_l}}_{\text{consumption externality}} - \underbrace{\frac{\partial c_l}{\partial N_l} Q_{jl}(\mathbf{L},\mathbf{p})}_{\text{production externality}} + \frac{\partial p_{jl}}{\partial N_l} Q_{jl}(\mathbf{L},\mathbf{p}) \tag{13}$$

The sign of $\frac{\partial Q_{jl}}{\partial N_l}$ depends on the relative strength of the market-size and market-share effects. If the market-size effect dominates, firms face higher demand in denser locations and thus have a demand-side incentive to agglomerate. The sign of $\frac{\partial c_l}{\partial N_l}$ depends on the effect of N_l on productivity A_l: higher productivity in denser locations lowers marginal costs, creating a supply-side incentive to agglomerate.[26] These effects are either mitigated or enhanced by the impact of N_l on equilibrium prices.

Equilibrium Given location characteristics $\{w_l, r_l, a_l\}$, the joint distribution of customer types and origins $G(q_i, z_i)$, and firm characteristics $\{x_j, \xi_j, z_j\}$, a spatial equilibrium is defined as the set of consumer location and product choices $\{P_{il}, P_{ij|l}\}$, and firm location, pricing, and input choices $\{\Psi_{jl}, \ell_l, h_l\}$ that satisfy the following:

[25] While commuting costs are typically modeled on the worker's side in the urban literature, they are rarely incorporated into the firm's location problem. Including them here is important for two reasons. First, they represent a substantial share of costs, amounting to roughly 22% of firms' monthly revenues. Second, 80.1% of the labor force in low-income countries is self-employed. For these individuals, the choice of workplace coincides with the choice of business location.

[26] Although I do not explicitly model land and labor markets, the data shows that rental costs and wages are higher in locations in denser locations. Hence, marginal costs may actually be higher in these locations.

1. Consumers choose location and product to maximize utility by equations (6) and (8);
2. Firms optimally choose inputs, prices, and location by equations (18), (10) and (12);
3. The goods market clears: $Q_{jl}(\mathbf{L}, \mathbf{p}) = A_l \ell^{\delta} h^{1-\delta}$.

As it is common in entry games, the equilibrium need not be unique. I defer the discussion of uniqueness to Section 6..

5 Estimation

I structurally estimate the model in three steps. First, taking firm location and production choices as given, I follow the approach of Goolsbee and Petrin (2004) to estimate the demand parameters $\{\alpha, \beta, \sigma, \theta, \tau_1, \mu, \gamma_l\}$, using a combination of transaction and mystery shoppers data. Second, given the estimated demand parameters, I use data on wages and rents to simulate firms' optimal choices of land and labor, and estimate firms' production function and production externality parameters $\{\delta, T_1, T_2\}$ via simulated method of moments. Finally, I apply Rust's Nested Fixed Point algorithm (Rust, 1987) to a static setting to estimate the commuting cost parameter $\{\tau_2\}$ in the entry game.

Model simplification To bring the model to the data, I introduce a few simplifying assumptions. I assume that there are two types of consumer in the economy: final consumers and retailers. These types differ in their price elasticities (α), preferences for quality (β), and outside options.[27] They also differ in the quantities they purchase per transaction: final consumers purchase one unit,[28] while retailers purchase ten units, the median transaction size for this type of customers in the data. Within each type, consumers also vary in their parish of origin.

Firms are heterogeneous in both quality and the owner's origin location. Product quality is measured using a score constructed via factor analysis, combining three dimensions: (i) the quality score assigned by the expert tailor, (ii) timely delivery in the mystery shoppers exercise, and (iii) the average price of fabrics typically used in production. Firms scoring above the median are classified as *high quality*, and those scoring below as *low quality*.[29]

A key aspect of the estimation is the definition of locations. I assume that a location corresponds to a parish, a choice I justify in Appendix B.1. To keep the estimation tractable

[27] See Appendix B.3.1 for details about the data used to compute outside options

[28] 59% of final customers in the transaction data buy one unit of good. 94% buy less than five units.

[29] I focus on the quality dimension for two reasons. First, quality is the most important characteristic customers consider when searching: 87% of customers cite product quality, compared to 58% who mention customer care and 40% who mention timely delivery. Second, the data show a strong correlation between prices and product quality (see Table 4). Including quality in the demand estimation is therefore crucial for addressing potential price endogeneity concerns.

and limit the number of potential equilibria, I restrict firm owners' location choice to either (i) establishing the firm in the parish where they reside, or (ii) locating in the core, which merges the three nearby parishes with the highest concentration of garment firms in the city center.[30]

5.1 First Step: Demand Parameters

I estimate demand using the two-step approach proposed by Goolsbee and Petrin (2004). In the first step, I use the transaction data to construct the following weighted log-likelihood function:

$$\ln L(\boldsymbol{\theta_1}, \boldsymbol{\delta}|\mathbf{L,p}) = \sum_{i,j,l} w_j \times I_{ijl} \times \ln P_{ijl}(\mathbf{L}, \mathbf{p}, \boldsymbol{\delta}; \boldsymbol{\theta_1}) \tag{14}$$

where P_{ijl} is the probability that a consumer of type i purchases from firm j in location l, w_j are sampling weights, and I_{ijl} is an indicator equal to one if a type-i consumer purchases from firm j in location l in the data. This step identifies the nonlinear parameters $\boldsymbol{\theta_1} = \{\sigma, \theta, \tau_1, \mu, \gamma_l\}$ and the mean utilities δ_{ijl}. Because microdata are available, the mean utilities can be treated as firm–consumer type fixed effects and included directly in the likelihood (Berry et al., 2004). Following Goolsbee and Petrin (2004), I maximize the likelihood only over $\boldsymbol{\theta_1} = \{\sigma, \theta, \tau_1, \mu, \gamma_l\}$ and solve for the vector $\boldsymbol{\delta}(\boldsymbol{\theta_1})$ that equates observed and predicted market shares. In the second step, I estimate the linear parameters $\boldsymbol{\theta_2} = \{\alpha, \beta\}$ from the following regression: $\delta_{ijl} = q_i^{\hat{\theta}}(\beta \mathbf{x}_j + \xi_j - \alpha p_{jl} q_i^{1-\hat{\theta}})$. To address potential correlation between unobserved quality ξ_j and prices, I instrument for prices using a cost shifter. The identification of the price coefficient is discussed in detail in the following section.

5.1.1 Identification

In this section, I discuss the identification of the demand parameters. I begin with the mean utility parameters $\{\alpha, \beta\}$ and explain in detail how I address price endogeneity. I then turn to the identification of the parameters governing consumers' preference for variety $\{\sigma, \theta\}$, the search parameters $\{\tau_1, \mu\}$, and finally the fixed effect $\{\gamma_l\}$.

Mean utility parameters The mean utility parameters α and β are identified from variation in within-location market shares across firms with heterogeneous prices and product quality. To address potential endogeneity arising from correlation between prices and unobserved product quality ξ_j, I instrument for prices using the cost of a zipper, which is observed at the firm level. Zippers are widely used in garment production and, unlike fabric,

[30] This simplification is consistent with the data: proximity to home is the main reason owners cite for preferring the periphery, and among owners who locate in the periphery, 50% establish their business in their parish of residence.

are homogeneous in quality. Their price may nonetheless vary across firms within the same location due to differences in supplier choice, purchasing volumes, access to credit, and search intensity. Although zippers account for a small share of total costs, such variation is likely to capture broader heterogeneity in input costs that is plausibly uncorrelated with product quality. Table A6 reports the first-stage results from the instrumental variable regression using transaction-level data. The estimates show that zipper costs are a strong predictor of prices, with a Cragg–Donald Wald F-statistics of 15.39 for the excluded instruments.[31]

Preference for variety The parameter σ governs the dispersion of taste shocks within a location: the larger the σ, the more similar consumers' preferences for different varieties, and the fiercer the competition among firms in the same location. Its identification parallels that of the nesting coefficient in nested logit models. Importantly, σ is not separately identified from the variance of the search cost shock ω_{il}, which is therefore normalized to one (Ben-Akiva and Lerman, 1985). Identification of σ relies on variation in the share of final consumers purchasing products across locations with different numbers of firms. To illustrate this, assume without loss of generality that all firms have the same mean utility $\delta_{ijl} = \bar{\delta}$ and set τ_1 and μ equal to zero. Under these assumptions, the share of *final consumers* purchasing from firm j in location l is:

$$P_{ijl}^{f} = \frac{\left(N_l \exp\left(\frac{\bar{\delta}}{1-\sigma}\right)\right)^{1-\sigma}}{1 + \sum_{k=1}^{L}\left(N_k \exp\left(\frac{\bar{\delta}}{1-\sigma}\right)\right)^{1-\sigma}} \tag{15}$$

It is straightforward to see that σ is pinned down by variation in market shares across locations with different numbers of firms. To discuss the identification of θ, I maintain the same assumptions and consider the share of *retailers* purchasing from firm j in location l:

$$P_{ijl}^{r} = \frac{\left(N_l \exp\left(\frac{\bar{\delta}}{q_i^{\theta}(1-\sigma)}\right)\right)^{q_i^{\theta}(1-\sigma)-1}}{\exp(u_{i0}^{r}) + \sum_{k=1}^{L}\left(N_k \exp\left(\frac{\bar{\delta}}{q_i^{\theta}(1-\sigma)}\right)\right)^{q_i^{\theta}(1-\sigma)}} \tag{16}$$

Given σ, θ is identified from variation in the share of retailers purchasing from firms in locations with different numbers of businesses. Intuitively, θ captures how dispersed retailers' idiosyncratic preferences are relative to those of final consumers.

[31] While firms are modeled as heterogeneous only in quality, I include additional controls in the regression of δ_{ijl} on prices to account for other unobserved heterogeneity. In particular, I control for the type of garment purchased in the transaction.

Search costs parameters The identification of the search cost parameters τ_1 and μ[32] relies on variation in market shares across locations that have the same inclusive value $IV_{il} = \sum_{j=1}^{N_j} N_{jl} \exp(\frac{\delta_{ijl}}{q_i^{\theta}(1-\sigma)})$, which summarizes the price, quality and number of products sold in a location.

Holding inclusive values constant, μ is identified from variation in market shares across locations that host the same total number of firms but differ in firm density. This parameter governs the distribution of within-location search costs: all else equal, denser locations attract more customers by lowering search costs. The parish area provides a key source of variation for the identification of μ, as it determines the surface over which consumers search when shopping in a given location. In the baseline estimation, parish area is defined by administrative boundaries. To ensure robustness and exclude areas that customers would never visit, I also re-estimate the model defining the parish area as the convex hull of the firms located there. The estimated parameters are nearly identical, as shown in Table A7.

The availability of price data is crucial for separately identifying μ from production externalities. If denser locations only attracted customers because agglomeration lowered marginal costs—and therefore prices—then, conditional on prices, there should be no residual differences in market shares across locations with heterogeneous density. The residual correlation between market shares and density that remains after conditioning on prices identifies the consumption externalities.

The parameter τ_1 is identified from variation in the share of customers purchasing from locations that differ in distance from consumers.[33] Identification of τ_1 is therefore possible only because the data include information on buyers' origins. Finally, the fixed effect γ_l is identified from variation in market shares across locations with differing numbers of firms in other sectors.[34] This variation is not absorbed by the number of garment firms because different sectors are clustered in distinct parts of the city center, as shown in Figure 1.

The three sources of variation are summarized by the following expression, which shows the

[32]Search costs are parametrized as $\tau_1 g_1(||z_i - z_l||) = \tau_1 \log(||z_i - z_l||)$ and $\mu(\frac{N_l}{a_l}) = \log(1 + \exp(\mu \frac{N_l}{a_l}))$.

[33]Distance is measured as driving time between the centroids of the customer's and the firm's parish. For final consumers, I use commuting distance from the place of residence; for retailers, from the workplace. I assume travel costs are identical across consumer types conditional on distance (i.e. τ_1 is not a function of the customer type). This assumption is supported by the data: after controlling for distance and destination parish, there are no significant differences in transport mode or travel expenditures between final consumers and retailers. A limitation of the data is that customer location is missing for around 19% of transactions. For an additional 16%, only the district—not the parish—of origin is reported. In these cases, customer location is imputed. See Appendix B.2 for details of the imputation procedure.

[34]In the estimation, γ_l is an indicator equal to one for parishes with an above-median number of firms in other sectors, where the median is calculated across all 96 parishes of Kampala. The number of non-garment firms is computed using data from the 2010 Census of Business Establishments.

ratio of customers purchasing from two locations with the same inclusive value, but different density of garment firms, other sector firms, or distance to consumers:

$$\frac{P_{il}}{P_{ih}} = \frac{\exp\left(-\mu\left(\frac{N_l}{a_l}\right) - \tau_1 g_1(||z_i - z_l||) + \gamma_l\right)}{\exp\left(-\mu\left(\frac{N_h}{a_h}\right) - \tau_1 g_1(||z_i - z_h||) + \gamma_h\right)} \tag{17}$$

5.2 Second Step: Supply Parameters

I estimate the supply-side parameters via simulated method of moments, using data on rents, wages, labor, and land from the firm survey. These parameters include the production function parameter δ, and the production externality terms T_1 and T_2. Firms' optimal optimal choices of land and labor are given by:

$$h^*_{jl} = \frac{Q_{jl}}{\exp\left(T_1 + T_2\log(\frac{N_l}{a_l})\right)}\left(\frac{(1-\delta)w_l}{\delta r_l}\right)^{\delta}, \qquad \ell^*_{jl} = \frac{Q_{jl}}{\exp\left(T_1 + T_2\log(\frac{N_l}{a_l})\right)}\left(\frac{\delta r_l}{(1-\delta)w_l}\right)^{1-\delta} \tag{18}$$

Here, Q_{jl} is the demand for firm j in location l, simulated using the parameters estimated in the first step. Rents and wages are taken directly from the data. Although I do not explicitly model land and labor markets, I allow rents and wages to vary by location.[35] The estimation targets the following moments: the mean number of workers (including the owner), the mean size of business premises, and the mean worker-to-premises ratio across locations.

Identification Given r_l and w_l, the ratio of labor to land *within* locations identifies δ. The production externality parameters T_1 and T_2 are identified by the levels of land and labor employed by firms, conditional on demand. Specifically, T_1 is pinned down by the input choices of firms operating in locations with no other businesses, while T_2 is identified from variation in the levels of land and labor across locations with different firm densities. In the presence of production externalities, firms in denser locations should appear more productive and therefore employ fewer inputs.

As discussed earlier, the data does not allow me to disentangle the extent to which productivity differences across locations reflect true production externalities, unobserved sorting, or production amenities. However, under the assumptions that (i) the core has better amenities and (ii) attracts more productive firms, attributing all observed productivity differences to production externalities provides an upper bound for the strength of this channel. Both assumptions are reasonable, especially given evidence of complementarities between production externalities and firm productivity (Combes et al., 2012).

[35] Rents and wages are held fixed in the counterfactuals.

5.3 Third Step: Location Parameters

The estimation of the commuting parameter τ_2 follows the Nested Fixed Point algorithm of Rust (1987).[36] A fixed point of the NFXP is a pair $\{\tau_2^*, \mathbf{L}^*\}$ that satisfies:

(i) $\tau_2^* = \arg\max_{\boldsymbol{\theta}_3} \sum_j \sum_l \ln \Psi_{jl}(\mathbf{L}^*, \boldsymbol{\tau_2}) I_{lj}$

(ii) $\mathbf{L}^* = \Psi(\mathbf{L}^*, \tau_2^*)$

where I_{lj} is an indicator for firm j locating in l, and $\Psi(\mathbf{L}^*, \tau_2)$ is defined in equation (12). The estimation routine begins by solving the fixed-point mapping in equation (12) at an initial guess of the parameter τ_2. To obtain firm j's best response function $\Psi_j(\mathbf{L}, \tau_2)$, I compute variable profits $\pi_{jl}(\mathbf{L}, \mathbf{P})$ for all firms, which requires solving for the Nash-Bertrand equilibrium of the pricing game. Once the fixed point probabilities are obtained, they are used to construct the log-likelihood $\mathcal{L}(\mathbf{L}^0) = \sum_j \sum_l \ln I_{lj} \Psi_{jl}(\mathbf{L}^0, \tau_2)$, which is maximized with respect to τ_2. This procedure is repeated until both probabilities and parameters converge.

For this step, I consider garment firms in the entire city of Kampala, along with 34 additional parishes just outside the city perimeter, where some of the owners operating in the core reside. Firm owners' origin location is taken exogenously from the data.[37] As previously discussed, I assume that owners can only choose to locate their business either in their parish of residence or in the core. This restriction reduces the number of potential choices from 96 parishes to 2, simplifying the computation of a spatial equilibrium with heterogeneous firms, customers, and locations. It also reduces the number of potential equilibria by ensuring that firms coordinate to agglomerate in one location—the core.

5.4 Structural Results

Table 4 presents the estimates of the demand, supply, and location parameters. This section discusses the magnitude of the estimated parameters and the fit of the model.

Demand parameters Panel A of Table 4 reports the demand parameter estimates, with Table A6 in the appendix showing the corresponding first stage results for the estimation of the price coefficient. The estimates have the expected signs: the price coefficient is negative,

[36] The location parameter λ is not separately identified from τ_2, and is therefore normalized to be equal to the average firm variable profits. I re-estimate the location parameters for values of λ equal to 0.75 and 1.25 times the average firm variable profits. Table A8 reports the corresponding estimates of τ_2 and the resulting equilibria of the entry game. I select the baseline value of λ based on the trade-off between goodness of fit and the intent to minimize the role of idiosyncratic preferences.

[37] With the demand and supply parameters at hand, the only data required to compute firms' best responses in the entry game ($\Psi(\mathbf{L}, \tau_2)$) are the number garment firms operating in a parish, the number of firm owners born in a parish, factor prices and productivity. More detail about how this data is computed for parishes outside the sample is provided in Appendix B.3.2.

with an implied average price elasticity of -3.23, while the quality coefficient is positive and of similar magnitude for retailers and final customers. Retailers' demand is more than twice as price elastic as that of final consumers, with implied elasticities of -3.77 and -1.37 respectively. This finding is consistent with Grant and Startz (2022), who show that elasticities increase moving upstream in the supply chain.

Table 4: Estimated parameters

	Parameter	Estimate	Std Error
PANEL A: Demand			
Price final consumers	α_f	-0.107	(0.021)
Price retailers	α_r	-0.071	(0.028)
Quality final consumers	β_f	0.284	(0.057)
Quality retailers	β_r	0.289	(0.079)
Taste shocks correlation	σ	0.578	(0.033)
Quantity multiplier	θ	0.379	(0.035)
Travel cost	τ_1	1.399	(0.040)
Within location search cost	μ	-0.0016	(0.0003)
Other-sector firms FE	γ_l	0.361	(0.087)
PANEL B: Supply			
Labor share	δ	0.665	(0.017)
Baseline productivity	T_1	0.745	(0.490)
Production externalities	T_2	0.316	(0.116)
PANEL C: Location			
Commuting cost	τ_2	8.009	(1.515)

Note: Table 4 shows point estimates and standard errors for the model parameters. Standard errors are bootstrapped using 100 bootstrapped samples.

The parameters σ and θ govern preferences for variety among final consumers and retailers. A sufficient condition for retailers to have a stronger preference for variety than final consumers is that $q_{ret}^{\theta} > \frac{1}{1-\sigma}$. This condition is satisfied given the estimates, indicating that, all else equal, retailers are more inclined to search in locations with a larger number of firms.

The travel cost parameter τ_1 is positive, consistent with consumers disliking distance. The implied average elasticity of consumption with respect to travel time is 1.23, slightly smaller than the elasticity estimated by Miyauchi et al. (2021) for Tokyo (1.99 to 2.92). This difference likely reflects the large share of retail customers in my sample, who are less sensitive to distance due to economies of scale in transport. On average, transport costs account for 9.6% of the transaction value, closely matching the 8.6% observed in the survey data.

This share varies sharply by consumer type—18.4% for final consumers versus 1.7% for retailers—again consistent with economies of scale in transport. The firm-specific search cost parameter μ is negative, indicating that within-location search is less costly denser areas. The estimates imply that search costs in the core are just 15.3% of those in the periphery. Finally, γ_l is positive, suggesting that consumers prefer to search in locations that also have a high concentration of non-garment firms.

To assess model fit, Panel A of Figure A10 compares the estimated and predicted share of customers purchasing products in each parish. The model closely tracks the data, indicating it captures the key determinants of location-level demand. Panel B examines the share of customers purchasing from a given firm within a location. This tests whether variation in prices and quality explains within-location allocation of customers to firms. The estimated shares broadly follow the data, though with considerable noise, which is expected given the limited firm heterogeneity in the model.

In Appendix C.1, I relax the assumption that the model is static by introducing a second period in which customers may switch firms (within or across locations) after observing taste shocks from the first period. Assuming customers draw the same firm-specific search costs and idiosyncratic taste shocks across periods, I find that 90% return to the same firm as in the first period. This suggests that the cross-sectional data used for estimation largely reflects customers' initial search process, consistent the model set-up.

Supply parameters Panel B of Table 4 reports the supply-side estimates, while Table A8 compares the data and simulated moments to assess model fit. The estimate of T_2 is positive, consistent with firm productivity increasing with density. The implied agglomeration elasticity of 0.316 is slightly larger than existing estimates but close to the value reported by Tsivanidis (2022) for Bogota (0.242). This is consistent with the estimate serving as an upper bound on the strength of production externalities.

Location parameters The last panel of Table 4 reports the estimated commuting cost parameter. The implied semi-elasticity with respect to distance (in minutes) is -0.0278, which is in line with estimates for Bogota (Tsivanidis, 2022), Mexico City (Zárate, 2020), and Medellín (Khanna et al., 2023), where values range between -0.02 and -0.05. In the equilibrium of the entry game associated with these parameters, 43.2% of low-quality firms and 49.6% of high-quality firms choose to locate in the core. These figures closely match the observed data shares (44.8% and 50.4%, respectively).

6 Counterfactuals

In this section, I construct three sets of counterfactuals. First, I examine how firms' spatial dispersion, profits, and consumer welfare would change in the absence of within-location search costs. Second, I assess the contribution of different channels—information frictions, production externalities, geographical centrality, and proximity to firms in other sectors—to the agglomeration of firms in the core. Third, I analyze the effects of two policies aimed at decongesting central Kampala: the introduction of an e-commerce platform and the eviction of firms from the core. I evaluate these policies under two alternative model specifications: the baseline model presented in Section 4 and a discrete choice demand model in which consumers are perfectly informed.

6.1 Eliminating Within-Location Search Costs

In the first counterfactual, I study the effect of eliminating search costs within locations. This exercise is not intended to simulate a real-world policy, but rather to quantify the impact of consumer information frictions on firm location and welfare. In this scenario, I assume that, upon visiting a location, consumers can observe all products sold there at no cost, though purchases still require in-person visits. I then re-compute variable profits under the new demand system and solve for the Nash equilibrium of the location game by searching for the fixed point of the system of best response functions in equation (12). Since the model can feature multiple equilibria, I search for fixed points starting from different spatial distributions: (i) all firms in the core; (ii) all firms in the periphery; (iii) one type of firms (e.g., high-quality) in the core and the other in the periphery. All searches converge to the same equilibrium, which suggests that it is unique.

Results are reported in Table 5. Column 1 shows the share of firms in the core, average prices, profits, and consumer welfare[38] in the baseline scenario with within-location search costs. Column 2 shows the same statistics for the no-search-cost counterfactual.

Eliminating search costs reduces the share of firms in the core by 18.8%. This aggregate change masks a substantial shift in firm composition across locations. In the baseline, half of all high-quality firms choose to locate in the core. This is because, as discussed in Section 4.2, they benefit the most from agglomeration under imperfect consumer information. Once search costs are removed, the share of high-quality firms in the core falls by 22.5%, compared to a 14.1% decline for low-quality firms. This reallocation leads to a 36.6% reduction in the share of total sales in the core.

[38] As standard in discrete choice models, consumer welfare is computed as the expected maximum value of utility divided by the price coefficient: $E_{\varepsilon,\omega}(\max_j U^q_{ijl})/\alpha$

Average firm profits rise by 3.81% when search costs are eliminated. This increase is driven by two forces. First, owners' commuting costs fall as more firms relocate closer to home. Second, the absence of search costs expands demand by inducing some customers who previously chose the outside option to search for garment products. This latter effect illustrates how information frictions can act as a demand-side constraint on firm growth. Prices, however, remain largely unchanged due to two offsetting forces. At the extensive margin, higher demand incentivizes firms to raise prices, while at the intensive margin, intensified competition from lower search costs pushes firms to reduce markups (Murry and Zhou, 2020; Moraga-González et al., 2023).

Table 5: Profits and welfare with no search costs

	Baseline	**No Search Costs**
Share of firms in core	0.466	0.378
		-18.8%
Share of high-quality in core	0.496	0.384
		-22.5%
Share of low-quality in core	0.432	0.371
		-14.1%
Share of sales in core	0.482	0.306
		-36.6%
Average price	9.002	9.125
		1.36%
Average profits	96.49	100.2
		3.81%
Average consumer welfare	45.94	50.31
		9.51%

Notes: Table 5 shows firm location, prices, profits and consumer welfare at the baseline equilibrium and in the counterfactual with no within-location search costs: $\mu(\frac{N_l}{a_l}) = 0$.

High-quality firms experience slightly larger profit gains (4.0%) compared to low-quality firms (3.6%). This difference arises because customers can now observe the full range of products available in a location, rather than only the limited subset reached under costly search. This disproportionately benefits higher-quality firms, which may have been overlooked when search costs restricted the set of visited firms. At the same time, reductions in commuting costs have a relatively larger effect on low-quality firms' profits, offsetting some of the effect of search costs alone. For consumers, the combination of costless search and lower average transport costs from firms relocating in the periphery result in a 9.5% increase in welfare.[39]

[39]In Appendix C.2, I extend the model to allow for firm entry and exit. Eliminating within-location

6.2 Decomposing Agglomeration

The model features four drivers of firm agglomeration in the core: consumer information frictions, production externalities, geographical centrality, and proximity to other-sector firms. In the second set of counterfactuals, I shut down these channels one by one to assess the contribution of each to the observed concentration of firms. The results are presented in Table 6.

Table 6: Firm agglomeration decomposition

	Change in parameters	**Change in share of firms in core**
Search frictions	$\mu(\frac{N_l}{a_l}) = 0$	18.8%
Production externalities	$T_2 = 0$	25.9%
Core geographical centrality	$\tau_1 = 0$	8.16%
Other-sector firms proximity	$\gamma_l = 0$	11.6%

Notes: Table 6 shows the change in the share of firms located in the core relative to the baseline equilibrium for each of the counterfactual scenarios listed in the table. The share of firms in the core across the different counterfactuals does not add up to 100% due to complementarities between the different channels and the presence of idiosyncratic location preferences, which lead some owners to always choose the same location.

Production externalities I begin with production externalities, the primary driver of agglomeration in most quantitative urban models. This counterfactual benchmarks the importance of this mechanism relative to consumption externalities from information frictions—the core addition of this paper to the quantitative spatial literature.

To quantify the role of production externalities, I recompute the equilibrium setting $T_2 = 0$. Under this scenario, 34.5% of firms choose to locate in the core, compared to 46.6% in the baseline and 37.8% in the counterfactual with no search costs. In other words, production externalities account for up to 25.9% of the observed clustering of garment firms in Kampala core. Since this estimate is an upper bound, the result suggests that consumption and production externalities are of comparable quantitative importance in this setting.

Centrality and proximity to other-sector firms The remaining two channels are the geographical centrality of the core and proximity to non-garment firms. The first channel resembles the classic mechanism in Hotelling (1931): although distant from each individual

search costs then leads to the entry of 2.8% more low-quality firms and 4.4% more high-quality firms. This moderates the profit effects—2.6% and 3.4% increases, respectively—while raising consumer welfare further (10.1%) through additional product variety and a higher share of high-quality firms.

parish, the core's central location makes it easily accessible from all parts of Kampala.[40] To quantify the strength of this channel, I recompute firms' location choices in the absence of consumers' transport costs by setting $\tau_1 = 0$. This reduces agglomeration by 8.16%. Proximity to other-sector firms, captured by the fixed effect γ_l, accounts for an additional 11.6% of agglomeration. This mechanism may reflect trip chaining, proximity to consumers' workplaces, or cross-sector production spillovers. However, in line with most sorting in this setting occurring within the same sector, this channel plays a secondary role.

6.3 Decongestion Policies

Firm agglomeration in the core comes at a high cost. In Kampala, travel time is estimated to account for 13.5% of the city GDP, with an additional 4.2% attributed to congestion (Baertsch, 2020). These figures have prompted the city administration to consider policies aimed at relocating informal businesses outside the city center. I simulate two such policies: the creation of an e-commerce platform where small vendors can sell their products, and the eviction of a share of the firms from the core. The distinction between these policies is crucial: the e-commerce platform targets the root cause of the inefficiency—information frictions—while firm evictions address only its symptom—agglomeration—without tackling the underlying cause. I compare their effects on firm location, profits, and consumer welfare under two different model specifications: (i) the baseline model presented in Section 4, and (ii) a discrete choice demand model that does not feature information frictions. Comparing policy outcomes across these two frameworks highlights how omitting information frictions from spatial models can lead to biased policy conclusions.

6.3.1 Policy Counterfactuals *with* Information Frictions

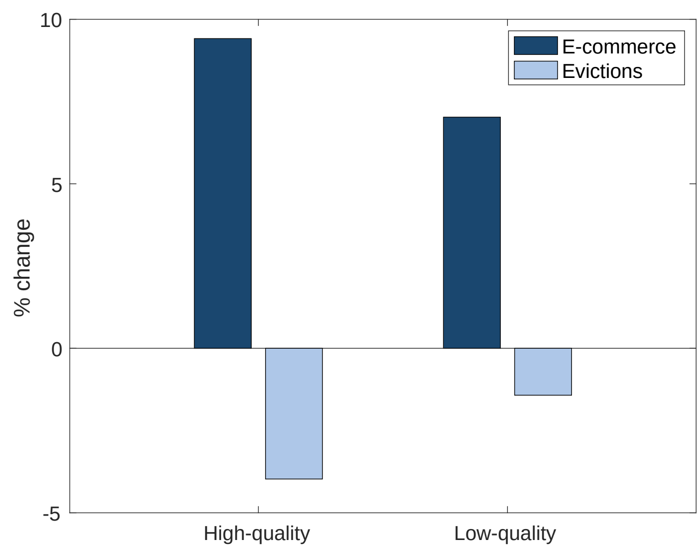
E-commerce platform In the e-commerce counterfactual, consumers may choose to purchase products online or continue traveling to firms in person. To use the platform, they pay a fixed cost to access the internet and a 20% commission fee on purchases.[41] Once on the platform, customers can observe all product varieties at no search cost and pay a flat delivery fee to receive products from any location in the city. This reduces two key drivers of agglomeration: search costs from information frictions and the geographical centrality of the core.

[40] This channel is reinforced by the city's radial road network, which the model incorporates by measuring distance as average driving time between locations.

[41] The cost of accessing internet is set equal to 1.11$, the average price of 1GB of internet in Uganda (source: cable.com.uk). The 20% fee reflects commissions charged by Jumia, Uganda's largest e-commerce platform, in the fashion category. I assume complete pass-through of this fee to consumers. Delivery costs are also calibrated from Jumia ($1.58) and are homogeneous across all locations in Kampala.

The results, reported in Table A10, show that the platform reduces the share of firms concentrated in the core by 32.3%. As in the no-search-cost counterfactual, this decline is primarily driven by high-quality firms relocating outside the core, leading to a 57.5% drop in the share of sales in this area. Lower transport and search costs raise consumer welfare and overall demand. Combined with reduced commuting costs for firm owner, this increases average profits by 8.4%. Figure 4 shows that high-quality firms benefit the most, with a 9.3% rise in profits compared to 7.2% for low-quality firms. This differential arises because the platform allows customers to compare all products available in the market, expanding their choice set and shifting some demand toward higher-quality firms that were previously less visible.

Figure 4: Impact on firm profits: e-commerce vs. caps



Notes: Figure 4 plots percentage changes in the profits of high- and low-quality firms relative to the baseline scenario. Dark blue bars represent the e-commerce platform counterfactual, while light blue bars represent the eviction policy.

Firm evictions I next compare the e-commerce platform with a policy that caps the number of firms allowed in the core by evicting all unauthorized businesses. For comparability, I simulate caps that produce the same spatial dispersion as the e-commerce platform.

Figure 4 shows that evictions unequivocally reduce average profits. This occurs because forcing firms to relocate outside the core prevents them from benefiting from consumption externalities generated by information frictions.[42] At the same level of spatial dispersion that is induced by the e-commerce platform, average profits fall by 10.2%. The losses are

[42]This is consistent with Bassi et al. (2022a), who show that firms in Uganda benefit from locating near major roads through improved visibility. They find that randomly dispersing firms reduces profits substantially, with the losses outweighing potential worker welfare gains from reduced pollution. Similarly Gechter and Kala (2025) find that the relocation of small firms to industrial areas outside New Delhi increased their probability of exiting the market.

more pronounced among high-quality firms, whose profits decline by 11.9% (compared to 7.7% for low-quality firms), reflecting their greater reliance on visibility. These outcomes are the reverse of those generated by the e-commerce platform, underscoring the importance of addressing the root cause of inefficiency rather than its symptoms.

6.3.2 Policy Counterfactuals *without* Information Frictions

To assess the importance of incorporating information frictions in the spatial model, I re-estimate the model under the assumption that consumers face no such frictions. I then simulate the impact of the same decongestion policies—e-commerce and evictions—under this alternative specification.

Table A13 reports the estimated parameters from the model without information frictions. Three key differences emerge relative to the baseline. First, σ is slightly larger. This is intuitive: without frictions, more of the variation in customer shares across locations with different numbers of firms is attributed consumers' preference for variety. Second, consumers' transport costs τ_1 are substantially lower. The implied elasticity of consumption with respect to travel is 0.728, well below estimates in the literature. This lower estimate arises because, in the absence of information frictions, agglomeration becomes less attractive. To rationalize why many customers still search in the core, transport costs must therefore be lower. Finally, for the same reason, the impact of proximity to other-sector firms γ_l is estimated to be negative, in contrast to the findings of Miyauchi et al. (2021) and Oh and Seo (2022).

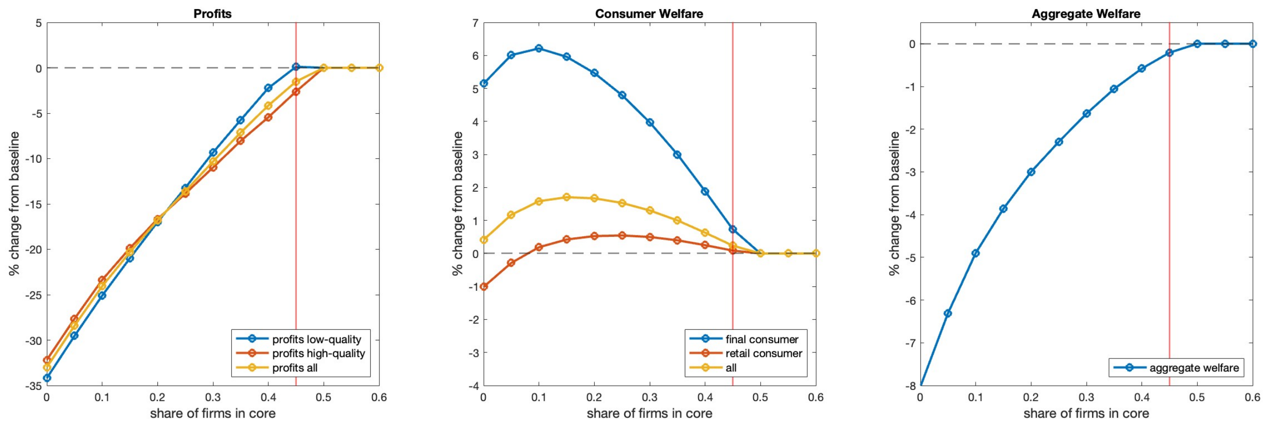
E-commerce platform I replicate the e-commerce counterfactual under the alternative model without information frictions. In this set-up, the only effect of the e-commerce platform is to reduce and equalize transport costs across locations.[43] The impact of the platform on firm location is negligible, leading, if anything, to a slight increase in the share of firms in the core (48% vs. 49.5%). Two factors explain this. First, since consumers can observe product varieties without the platform, its creation does not alter firm incentives to agglomerate for visibility. Second, because transport costs are estimated to be much lower, equalizing them across locations has only modest consequences. The estimated gains in consumer welfare and firm profits remain positive, but are much smaller than in the baseline model, at 3.3% (vs. 30.4%) and 1.1% (vs. 8.5%), respectively. In addition, the counterfactual predicts similar effects of the platform on low- and high-quality firms, with profits increases of 0.9% and 1.3%, respectively.

[43] Consumers still pay the same fixed cost to access the internet and a 20% commission fee on purchases

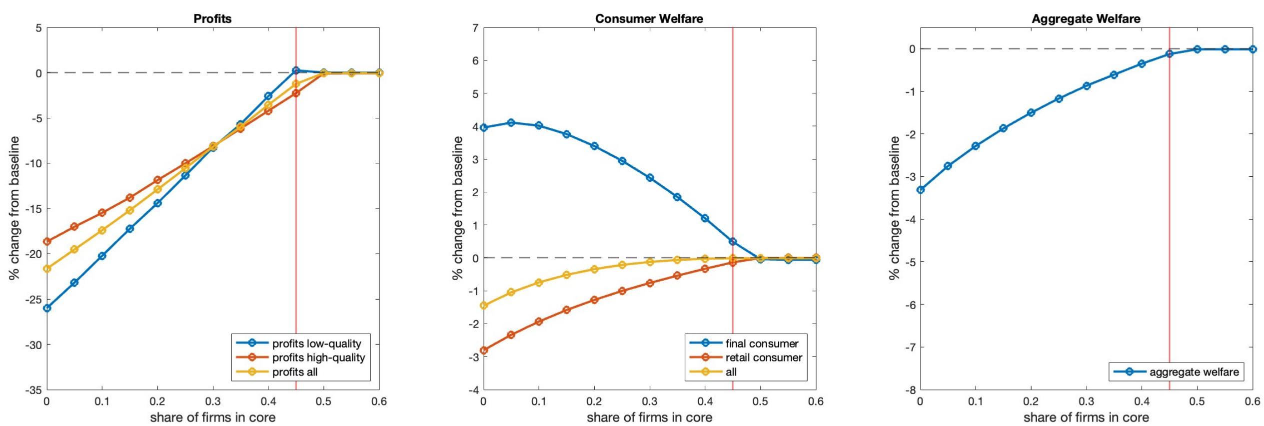
Firm evictions To better understand how omitting information frictions affects policy counterfactuals, Figure 5 shows the estimated impact of firm evictions on profits, consumer welfare, and overall welfare for the baseline model and the alternative model without information frictions. Specifically, the figure reports the percentage change in each outcome when only the share of firms indicated on the x-axis is permitted to remain in the core, with any excess firms randomly evicted.[44]

Figure 5: Impact of evictions

Panel A: *With* search costs



Panel B: *Without* search costs



Notes: Figure 5 shows percentage changes in profits, consumer welfare, and aggregate welfare when only the share of firms indicated on the x-axis is allowed to operate in the core, relative to the baseline scenario with no restrictions. Panel A shows predicted changes from a model with within-location search costs. Panel B shows predicted changes from a model without such costs. In each panel, the first figure shows the effects on firm profits, separately for low- and high-quality firms. The second figure shows the effects on consumer welfare, separately for final consumers and retailers. The yellow lines shows average impacts.

In both models, evictions reduce firm profits. However, the model without information frictions systematically underestimates the negative effects of the policy. For example, re-

[44] I assume the same share of high- and low-quality firms are allowed to operate in the core.

stricting core activity to 25% of firms—about half of the current share—yields an 10.6% decline in profits in the model without frictions, compared to a 13.6% loss in the baseline. Profit losses are especially underestimated for high-quality businesses (by 28%, compared to 14% for low-quality firms). This occurs because, in the frictionless model, location does not influence a firm's likelihood of entering consumers' choice set, and agglomeration therefore does not disproportionately benefit high-quality firms.

Turning to consumer welfare, two patterns emerge. First, final consumers are harmed by agglomeration in both counterfactuals. Firms cluster primarily to attract retailers, who buy in bulk and face low per-unit transport costs. Final consumers, by contrast, would be better off with more firms in the periphery, where they could access a broader variety of products and, in the baseline model, face lower within-location search costs.

Second, the impact on retailers differs across the two models. In the model without information frictions, the only sources of agglomeration are production externalities and, to a lesser extent, core centrality. Evictions raise marginal costs and therefore prices by limiting the former (see Figure A11), and this effect dominates the benefit of lower transport costs, leading retailer welfare to fall. By contrast, in the baseline model, a moderate level of evictions can actually benefit retailers. The reason is a non-linear relationship between firm density and within-location search costs: the marginal benefit of an additional firm on search costs declines as density rises. Relocating firms to the periphery can therefore substantially lower search costs there (while leaving search costs in the core largely unchanged), improving consumer welfare. Because firms do not internalize these externalities, they tend to over-agglomerate relative to the consumer-optimal level.[45]

Both models predict overall welfare losses from evictions. Yet because the frictionless model underestimates the impact on profits, it also underestimates total welfare losses. For instance, when only 25% of firms are allowed to remain in the core, it predicts a welfare decline of just 1.2%, almost half of the 2.3% reduction implied by the baseline model.[46]

7 Conclusion

Several studies have documented key differences in the spatial organization of cities across the development spectrum. Despite this, most spatial model developed to study these cities

[45] Formally, the derivative of aggregate consumer welfare with respect to N_{jl} is $\sum_i N(i)\big(q_i^\theta(1-\sigma)P_{ijl} - P_{il}\mu'\big(\frac{N_l}{a_l}\big)\big)$. The second derivative is $\sum_i N(i)\big(q_i^\theta(1-\sigma)\frac{\partial P_{ijl}}{\partial N_{jl}} - \frac{\partial P_{il}}{\partial N_{jl}}\mu'\big(\frac{N_l}{a_l}\big) - P_{il}\mu''\big(\frac{N_l}{a_l}\big)\big)$. In the presence of strong consumption externalities, the first two terms are positive, but the last is negative, implying diminishing marginal benefits of density.

[46] The difference in estimated profit losses outweighs the larger gains in consumer welfare in the baseline relative to the frictionless model.

do not incorporate markets frictions that are common in the developing world. This paper takes a step in that direction by examining the role of consumers' limited access to product information on firms' incentives to agglomerate. To do that, I collect unique firm and customer data and use them to estimate a quantitative model of consumer search and firm location.

There are four takeaways from this study. First, information frictions are a key driver of firm agglomeration, playing a role that is quantitatively similar to that of widely studied production externalities. Second, high search costs associated with information frictions reduce demand and firm profits, constituting a potential constraint to firm growth. Third, policies focused solely on reducing agglomeration can disproportionately penalize high-quality firms by increasing consumers' costs of gathering information. Fourth, omitting information frictions from spatial models introduces significant biases in the estimates of the welfare effects of urban policies.

The paper identifies two potential areas for future research. First, it documents the interplay between firm agglomeration and the presence of bulk buyers in the market. In particular, the results from the eviction counterfactual highlights how agglomeration potentially harms final consumers. While the choices of firms and intermediaries have been separately studied in this paper and in recent work by Grant and Startz (2022), jointly examining the linkages between firms' location and intermediaries' entry decisions could advance our understanding of supply chain formation in low-income settings. Second, the paper highlights how incorporating market frictions that characterize the developing world into spatial models is essential for predicting the effects of urban policies. As Bryan et al. (2025) recently emphasized, identifying and integrating such frictions into models of developing-country cities is a crucial step toward advancing our understanding of how how cities in low-income countries evolve and respond to policy.

References

Agarwal, S., J. B. Jensen, and F. Monte (2017) "Consumer mobility and the local structure of consumption industries," *NBER Working Paper N. 23616.*

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015) "The economics of density: Evidence from the Berlin Wall," *Econometrica*, Vol. 83, pp. 2127–2189.

Aker, J. C. (2010) "Information from markets near and far: Mobile phones and agricultural markets in Niger," *American Economic Journal: Applied Economics*, Vol. 2, pp. 46–59.

Allen, T. (2014) "Information frictions in trade," *Econometrica*, Vol. 82, pp. 2041–2083.

Allen, T., C. Arkolakis, and X. Li (2015) "Optimal city structure," *Yale, mimeo.*

Anderson, S. P., A. De Palma, and J.-F. Thisse (1992) *Discrete choice theory of product differentiation*: MIT press.

Armstrong, M. (2017) "Ordered consumer search," *Journal of the European Economic Association*, Vol. 15, pp. 989–1024.

Atkin, D., A. K. Khandelwal, and A. Osman (2017) "Exporting and firm performance: Evidence from a randomized experiment," *The Quarterly Journal of Economics*, Vol. 132, pp. 551–615.

Baertsch, L. (2020) "Quantifying the economic benefits of public transportation in Kampala," *IGC Policy Brief*.

Balboni, C., G. Bryan, M. Morten, and B. Siddiqi (2020) "Transportation, gentrification, and urban mobility: The inequality effects of place-based policies," *LSE, mimeo*.

Bandiera, O., A. Elsayed, A. Smurra, and C. Zipfel (2022) "Young adults and labor markets in Africa," *Journal of Economic Perspectives*, Vol. 36, pp. 81–100.

Baseler, T. (2023) "Hidden income and the perceived returns to migration," *American Economic Journal: Applied Economics*, Vol. 15, pp. 321–352.

Bassi, V., M. E. Kahn, N. L. Gracia, T. Porzio, and J. Sorin (2022a) "Searching for Customers, Finding Pollution," *NBER Working Paper N. 30536*.

Bassi, V., R. Muoio, T. Porzio, R. Sen, and E. Tugume (2022b) "Achieving scale collectively," *Econometrica*, Vol. 90, pp. 2937–2978.

Ben-Akiva, M. E. and S. R. Lerman (1985) *Discrete choice analysis: theory and application to travel demand*, Vol. 9: MIT press.

Bergquist, L. F., C. McIntosh, and M. Startz (2021) "Search cost, intermediation, and trade: Experimental evidence from Ugandan agricultural markets," *Yale, mimeo*.

Berry, S., J. Levinsohn, and A. Pakes (2004) "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of political Economy*, Vol. 112, pp. 68–105.

Berry, S. T. (1994) "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, pp. 242–262.

Bloom, N., A. Mahajan, D. McKenzie, and J. Roberts (2010) "Why do firms in developing countries have low productivity?" *American Economic Review*, Vol. 100, pp. 619–23.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014) "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," *Econometrica*, Vol. 82, pp. 1671–1748.

Bryan, G., E. Glaeser, and N. Tsivanidis (2020) "Cities in the developing world," *Annual Review of Economics*, Vol. 12, pp. 273–297.

Bryan, G. T., K. Frye, and M. Morten (2025) "Spatial economics for low-and middle-income countries," *NBER Working Paper N. 33453*.

Carrillo, P., D. Donaldson, D. Pomeranz, and M. Singhal (2023) "Misallocation in Firm Production: A Nonparametric Analysis Using Procurement Lotteries," *NBER Working Paper N. 31311*.

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (2017) "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States," *Journal of Urban Economics*, Vol. 98, pp. 17–49.

Choi, M., A. Y. Dai, and K. Kim (2018) "Consumer search and price competition," *Econometrica*, Vol. 86, pp. 1257–1281.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012) "The productivity advantages of large cities: Distinguishing agglomeration from firm selection," *Econometrica*, Vol. 80, pp. 2543–2594.

Combes, P.-P. and L. Gobillon (2015) "The empirics of agglomeration economies," in *Handbook of Regional and Urban Economics*, Vol. 5: Elsevier, pp. 247–348.

Couture, V. (2016) "Valuing the consumption benefits of urban density," *UC Berkeley, mimeo*.

Deffebach, P., D. Lagakos, Y. Miyauchi, and E. Yamada (2025) "The Spatial Distribution of Income in Cities: New Global Evidence and Theory," *Boston University, mimeo*.

Dingel, J. I. and F. Tintelnot (2020) "Spatial economics for granular settings," *NBER Working Paper N. 27287*.

Dudey, M. (1990) "Competition by choice: The effect of consumer search on firm location decisions," *American Economic Review*, pp. 1092–1104.

Duranton, G. and D. Puga (2004) "Micro-foundations of urban agglomeration economies," in *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2063–2117.

Fischer, J. H. and J. E. Harrington Jr (1996) "Product variety and firm agglomeration," *The RAND Journal of Economics*, pp. 281–309.

Gechter, M. and N. Kala (2025) "Firm presence, pollution, and agglomeration: Evidence from a randomized rnvironmental place-based policy," *NBER Working Paper N. 33707*.

Gechter, M. and N. Tsivanidis (2023) "Spatial spillovers from high-rise developments: Evidence from the Mumbai Mills," *mimeo, University of California Berkeley Haas.*

Gertler, P., M. Gonzalez-Navarro, R. Undurraga, and J. A. Urrego (2024) "In-situ Upgrading or Population Relocation? Direct Impacts and Spatial Spillovers of Slum Renewal Policies," *mimeo, University of California Berkeley.*

Goldberg, P. and T. Reed (2023) "Demand-Side Constraints in Development: The Role of Market Size, Trade, and (In) Equality," *Econometrica*, Vol. 91, pp. 1959–1962.

Goolsbee, A. and A. Petrin (2004) "The consumer gains from direct broadcast satellites and the competition with cable TV," *Econometrica*, Vol. 72, pp. 351–381.

Goyal, A. (2010) "Information, direct access to farmers, and rural market performance in central India," *American Economic Journal: Applied Economics*, Vol. 2, pp. 22–45.

Grant, M. and M. Startz (2022) "Cutting out the middleman: The structure of chains of intermediation," *NBER Working Paper N. 30109.*

Harari, M., M. Wong et al. (2018) "Slum upgrading and long-run urban development: Evidence from indonesia," *mimeo, Wharton.*

Hjort, J., V. Iyer, and G. De Rochambeau (2020) "Informational barriers to market access: experimental evidence from Liberian firms," *NBER Working Paper N. 27662.*

Hong, H. and M. Shum (2006) "Using price distributions to estimate search costs," *The RAND Journal of Economics*, Vol. 37, pp. 257–275.

Hortaçsu, A. and C. Syverson (2004) "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *The Quarterly Journal of Economics*, Vol. 119, pp. 403–456.

Hotelling, H. (1931) "The economics of exhaustible resources," *Journal of Political Economy*, Vol. 39, pp. 137–175.

Hsieh, C.-T. and B. A. Olken (2014) "The missing "missing middle"," *Journal of Economic Perspectives*, Vol. 28, pp. 89–108.

Jedwab, R., P. Loungani, and A. Yezer (2021) "Comparing cities in developed and developing countries: Population, land area, building height and crowding," *Regional Science and Urban Economics*, Vol. 86, p. 103609.

Jensen, R. (2007) "The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector," *The Quarterly Journal of Economics*, Vol. 122, pp. 879–924.

Jensen, R. and N. H. Miller (2018) "Market integration, demand, and the growth of firms: Evidence from a natural experiment in India," *American Economic Review*, Vol. 108, pp. 3583–3625.

KCCA, K. C. C. A. (2019) "Statistical abstract for Kampala City,"Technical report.

Khanna, G., A. Nyshadham, D. Ramos-Menchelli, J. A. Tamayo, and A. Tiew (2023) "Spatial mobility, economic opportunity, and crime," *HBS, mimeo.*

Konishi, H. (2005) "Concentration of competing retail stores," *Journal of Urban Economics*, Vol. 58, pp. 488–512.

Kreindler, G., A. Gaduh, T. Graff, R. Hanna, and B. A. Olken (2023) "Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta," *NBER Working Paper N. 31369*.

Lagakos, D. (2016) "Explaining cross-country productivity differences in retail trade," *Journal of Political Economy*, Vol. 124, pp. 579–620.

Le Barbanchon, T., R. Rathelot, and A. Roulet (2021) "Gender differences in job search: Trading off commute against wage," *The Quarterly Journal of Economics*, Vol. 136, pp. 381–426.

Michaels, G., D. Nigmatulina, F. Rauch, T. Regan, N. Baruah, and A. Dahlstrand (2021) "Planning ahead for better neighborhoods: Long-run evidence from Tanzania," *Journal of Political Economy*, Vol. 129, pp. 2112–2156.

Miyauchi, Y., K. Nakajima, and S. J. Redding (2021) "Consumption access and the spatial concentration of economic activity: evidence from smartphone data," *NBER Working Paper N. 28497*.

Mizuno, T. (2003) "On the existence of a unique price equilibrium for models of product differentiation," *International Journal of Industrial Organization*, Vol. 21, pp. 761–793.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018) "Commuting, migration, and local employment elasticities," *American Economic Review*, Vol. 108, pp. 3855–90.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2023) "Consumer search and prices in the automobile market," *The Review of Economic Studies*, Vol. 90, pp. 1394–1440.

Murry, C. and Y. Zhou (2020) "Consumer search and automobile dealer colocation," *Management Science*, Vol. 66, pp. 1909–1934.

Oh, R. and J. Seo (2022) "What Causes Agglomeration of Services? Theory and Evidence from Seoul," *Yale, mimeo.*

Owens, R., E. Rossi-Hansberg, and P.-D. Sarte (2020) "Rethinking detroit," *American Economic Journal: Economic Policy*, Vol. 12, pp. 258–305.

Pelnik, C. (2024) "Moving to Profitability? Alleviating Constraints on Microentrepreneur Location," *World Bank, mimeo.*

Redding, S. J. (2023) "Quantitative urban models: from theory to data," *Journal of Economic Perspectives*, Vol. 37, pp. 75–98.

Rosenthal-Kay, J. D. (2025) "Urban Costs Around the World," *University of Chicago, mimeo.*

Rosenthal, S. S. and W. C. Strange (2004) "Evidence on the nature and sources of agglomeration economies," in *Handbook of Regional and Urban Economics*, Vol. 4: Elsevier, pp. 2119–2171.

Rust, J. (1987) "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, pp. 999–1033.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012) "Testing models of consumer search using data on web browsing and purchasing behavior," *American Economic Review*, Vol. 102, pp. 2955–80.

Seim, K. (2006) "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, Vol. 37, pp. 619–640.

Stahl, K. (1982) "Differentiated products, consumer search, and locational oligopoly," *The Journal of Industrial Economics*, pp. 97–113.

Startz, M. (2021) "The value of face-to-face: Search and contracting problems in Nigerian trade," *Dartmouth, mimeo.*

Tsivanidis, N. (2022) "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio," *Berkeley Haas, mimeo.*

UN, U. N. (2018) "World Urbanization Prospects: The 2018 Revision."

Verhoogen, E. A. (2008) "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, Vol. 123, pp. 489–530.

Weitzman, M. (1979) "Optimal search for the best alternative," *Econometrica*, Vol. 47, pp. 641–654.

Wolinsky, A. (1983) "Retail trade concentration due to consumers' imperfect information," *The Bell Journal of Economics*, pp. 275–282.

World Bank (2021) "Decarbonizing Cities by Improving Public Transport and Managing Land Use and Traffic,"Technical report, World Bank.

Zárate, R. D. (2020) "Spatial misallocation, informality, and transit improvements: Evidence from Mexico City," *World Bank, mimeo.*

Supplemental Appendix

A Appendix A: Proofs and Derivations

A.1 Search costs distribution

Let F^w_{ijl} be the CDF of w_{ijl}, which can be computed as the distribution of the minimum of two independent random variables:[47]

$$F^w_{ijl}(z) = 1 - (1 - F^r_{ijl}(z))(1 - F_{ijl}(z)) \tag{A.1}$$

Using the expression for the reservation value $H_{ijl}(r_{ij}) = \delta_{ijl} + \int_{r_{ijl}}^{\infty}((1-\sigma)q^\theta z - r_{ijl})dF(z)$:

$$F^r_{ijl}(z) = \Pr(r_{ijl} < z) = \Pr[H_{ijl}(r_{ijl}) > H_{ijl}(z)] = \Pr[c_{ijl} > H_{ijl}(z)] = 1 - F^c_{ijl}(H_{ijl}(z)) \tag{A.2}$$

Plugging this expression into (A.1), I obtain:

$$F^w_{ijl}(z) = 1 - F^c_{ijl}(H_{ijl}(z))(1 - F_{ijl}(z)) \tag{A.3}$$

This expression provides an explicit relationship between the search cost distribution and the distribution of w. Following Moraga-Gonzalez et al. (2023), I show that, for a specific distribution of search costs, w is distributed as T1EV. To derive this distribution, I first show that consumer i reservation value for firm j in location l can be written as:

$$r_{ijl} = \delta_{ijl} + H_0^{-1}(c_{ijl}) \tag{A.4}$$

where $H_0(r) = \int_r^\infty (z - r)dF(\frac{z}{q^\theta(1-\sigma)})$. To show it, first notice that we can rewrite H_{ijl} from equation (4) as:

$$H_{ijl}(r) = \int_r^\infty (z - r)dF\left(\frac{z - \delta_{ijl}}{q^\theta(1-\sigma)}\right) \tag{A.5}$$

Using the change of variables $t = z - \delta_{ijl}$, I obtain:

$$H_{ijl}(r) = \int_{r-\delta_{ijl}}^\infty (t - (r - \delta_{ijl}))dF\left(\frac{t}{q^\theta(1-\sigma)}\right) = H_0(r - \delta_{ijl}) \tag{A.6}$$

Combining $H_{ijl}(r) = c_{ijl}$ and equation (A.6), we obtain $H_{ijl}(r) = H_0(r - \delta_{ijl}) = c_{ijl}$. We can

[47] r_{ijl} and u_{ijl} are independent conditional on δ_{ijl} because the former is a function of the search costs c_{ijl}, while the latter is a function of the taste shocks ε_{jl}, which are independently distributed.

invert this expression to get $r_{ijl} - \delta_{ijl} = H_0^{-1}(c_{ijl})$. I can now derive the expression for the CDF of w. First, I rewrite equation (A.3) as:

$$F_{ijl}^c(H_{ijl}(z)) = \frac{1 - F_{ijl}^w(z)}{1 - F_{ijl}(z)} \tag{A.7}$$

Using the change of variables $H_{ijl}(z) = c$, we can rewrite this distribution as:

$$F_{ijl}^c(c) = \frac{1 - F_{ijl}^w(H_{ijl}^{-1}(c))}{1 - F_{ijl}(H_{ijl}^{-1}(c))} = \frac{1 - F_{ijl}^w(\delta_{ijl} + H_0^{-1}(c))}{1 - F_{ijl}(\delta_{ijl} + H_0^{-1}(c))} = \frac{1 - F_{ijl}^w(\delta_{ijl} + H_0^{-1}(c))}{1 - F\big(\frac{H_0^{-1}(c)}{q^\theta(1-\sigma)}\big)} \tag{A.8}$$

Let $\mu(\frac{N_l}{a_l})$ be the location parameter of the search cost distribution, which depends on the density of firms in the location. To obtain a closed-form solution for buying probabilities, we derive an expression of the search cost distribution F_{ijl}^c that ensures w_{ijl} follows a Gumbel distribution with location parameter $\delta_{ijl} - \mu(\frac{N_l}{a_l})$ and scale parameter $q^\theta(1-\sigma)$:

$$F_{ijl}^w(z) = \exp\Bigg(-\exp\Bigg(\frac{-(z - (\delta_{ijl} - \mu(\frac{N_l}{a_l})))}{q^\theta(1-\sigma)}\Bigg)\Bigg) \tag{A.9}$$

Substituting this expression in equation (A.7), we obtain the Gumbel-preserving distribution of within-location search costs:

$$F_{ijl}^c(c) = \frac{1 - \exp\Bigg(-\exp\Bigg(\frac{-\Big(H_0^{-1}(c) + \mu\Big(\frac{N_l}{a_l}\Big)\Big)}{q_i^\theta(1-\sigma)}\Bigg)\Bigg)}{1 - \exp\Big(-\exp\Big(\frac{H_0^{-1}(c)}{q_i^\theta(1-\sigma)}\Big)\Big)} \tag{A.10}$$

A.2 Optimal prices

The FOC for firms' optimal prices are:

$$p_{jl}(\mathbf{L},\mathbf{p}) = c_l - \frac{Q_{ijl}(\mathbf{L},\mathbf{p})}{\partial Q_{ijl}(\mathbf{L},\mathbf{p})/\partial p_{jl}} = c_l - \frac{\int q_i P_{ijl}(\mathbf{L},\mathbf{p}) dG(q,z)}{\int q_i \frac{\partial P_{ijl}(\mathbf{L},\mathbf{p})}{\partial p_{jl}} dG(q,z)} \tag{A.11}$$

where P_{ijl} is given by equation (10). Omitting the arguments $(\mathbf{L},\mathbf{p})$ from now on, the derivative within the integral in the denominator is equal to:

$$\frac{\partial P_{ijl}}{\partial p_{jl}} = \frac{\partial P_{il}}{\partial p_{jl}} P_{ij|l} + P_{il} \frac{\partial P_{ij|l}}{\partial p_{jl}} \tag{A.12}$$

where, $\frac{\partial P_{il}}{\partial p_{jl}} = -\alpha q_i P_{ijl}(1-P_{il})$ and $\frac{\partial P_{ij|l}}{\partial p_{jl}} = -\frac{\alpha q_i^{1-\theta}}{1-\sigma} P_{ij|l}(1-P_{ij|l})$. Putting everything together, the expression above becomes:

$$\frac{\partial P_{ijl}}{\partial p_{jl}} = -\frac{\alpha}{1-\sigma} q_i P_{ijl} \left[q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta}) \right] \tag{A.13}$$

where the terms in square brackets is greater than zero. Plugging this into equation (A.13), the expression for optimal prices becomes:

$$p_{jl}^* = c_l + \frac{1-\sigma}{\alpha} \left(\frac{\int q_i P_{ijl} dG(q,z)}{\int q_i^2 P_{ijl}(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta}))dG(q,z)} \right) \tag{A.14}$$

A.3 Existence and Uniqueness of price equilibrium

I use a result from Mizuno (2003) to find the conditions that guarantee the existence and uniqueness of a Nash-Bertrand equilibrium in the model. Let $D_j(p_j|\mathbf{p}_{-j})$ be the demand for product j in a differentiated products setting, and let $C_j(Q_j)$ be the firm j cost function. Mizuno (2003) proves that the following five conditions are sufficient for the existence of a unique price equilibrium:

(i) $D_j(p_j|\mathbf{p}_{-j})$ is strictly positive and strictly decreasing in p_j on R^n,

(ii) $D_j(\mathbf{p}) = D_j(\mathbf{p} + k\mathbf{u}^n)$ for all k, where u is the n vector whose elements are all unity,

(iii) $D_j(p_j^H|\mathbf{p}_{-j}^H)D_j(p_j^L|\mathbf{p}_{-j}^L) \geq D_j(p_j^H|\mathbf{p}_{-j}^L)D_j(p_j^L|\mathbf{p}_{-j}^H)$ for $p_j^H \geq p_j^L$ and $\mathbf{p}_{-j}^H \geq \mathbf{p}_{-j}^L$,

(iv) $D_j(p_j|\mathbf{p}_{-j})$ is increasing in $\mathbf{p}_{-j}$ on R^n, OR, $C_j(Q_j) = c_j Q_j$, where $c_j \geq 0$

It is easy to see that the model satisfies conditions (i), (ii) and (iv). Condition (iii) can be re-written as:

$$\frac{D_j(p_j^H|\mathbf{p}_{-j}^H) - D_j(p_j^L|\mathbf{p}_{-j}^H)}{D_j(p_j^L|\mathbf{p}_{-j}^H)} \geq \frac{D_j(p_j^H|\mathbf{p}_{-j}^L) - D_j(p_j^L|\mathbf{p}_{-j}^L)}{D_j(p_j^L|\mathbf{p}_{-j}^L)} \tag{A.15}$$

which is satisfied if $\frac{\partial D_j(\mathbf{p})}{\partial p_j}/D_j(\mathbf{p})$ is increasing in p_{-j}, namely if:

$$\frac{\partial^2 D_j(\mathbf{p})}{\partial p_j \partial p_{-j}} \frac{1}{D_j(\mathbf{p})} - \frac{1}{D_j(\mathbf{p})^2} \frac{\partial D_j(\mathbf{p})}{\partial p_j} \frac{\partial D_j(\mathbf{p})}{\partial p_{-j}} \geq 0 \tag{A.16}$$

I focus again on the case of firms facing one type of consumers and show under what conditions inequality (A.30) is satisfied. With one type of consumers, the demand for firm j operating in location l is $P_{ijl}(\mathbf{L},\mathbf{p})$, where I omit the arguments $\mathbf{L},\mathbf{p}$ from now on. Equation

(A.30) becomes:

$$\frac{\partial^2 P_{ijl}}{\partial p_{jl}\partial p_{-j}} - \frac{1}{P_{ijl}}\frac{\partial P_{ijl}}{\partial p_{jl}}\frac{\partial P_{ijl}}{\partial p_{-j}} \geq 0 \tag{A.17}$$

Where the price derivatives are:

$$(i) \quad \frac{\partial P_{ijl}}{\partial p_{jl}} = -\frac{\alpha q_i}{1-\sigma} P_{ijl}\Big(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta})\Big) \tag{A.18}$$

$$(ii) \quad \frac{\partial P_{ijl}}{\partial p_{kl}} = -\frac{\alpha q_i}{1-\sigma} P_{ijl}P_{ik|l}((1-\sigma)(1-s_{il}) - q_i^{-\theta}) \qquad for\ k \neq j \tag{A.19}$$

$$(iii) \quad \frac{\partial P_{ijl}}{\partial p_{kh}} = \alpha q_i P_{ijl} P_{ikh} \qquad for\ k \neq j\ and\ h \neq l \tag{A.20}$$

Notice that, for firm j in location l an increase in the price charged by a different firm in a different location has a positive effect on demand (equation (A.34)). However, the impact of an increase in the price charged by a different firm in the same location can be negative. This is because such an increase in prices makes the location less attractive, reducing demand for all firms in the same location. This effect can outweigh the fact that, within the location, an increase in the price charged by other businesses makes a firm more attractive.

I study the sign (A.31) separately for changes in the prices charged by firms in the same and in different locations.

(i) Change in price by firms in the same location

I start with the former case. The cross-derivative in equation (A.31) becomes:

$$\begin{aligned}\frac{\partial^2 P_{ijl}}{\partial p_{jl}\partial p_{kl}} = &\left(\frac{\alpha q_i}{1-\sigma}\right)^2 P_{ijl}P_{ik|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta})\Big(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta})\Big) \\ &-\frac{\alpha q_i^{1-\theta}}{1-\sigma} P_{ijl}P_{ij|l}P_{ik|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta}) - \alpha q_i P_{ijl}P_{ik|l}(1-\sigma)(1-P_{il})\end{aligned} \tag{A.21}$$

The second term in the inequality is equal to:

$$\frac{1}{P_{ijl}}\frac{\partial P_{ijl}}{\partial p_{jl}}\frac{\partial P_{ijl}}{\partial p_{kl}} = \left(\frac{\alpha q_i}{1-\sigma}\right)^2 P_{ijl}P_{ik|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta})\Big(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta})\Big) \tag{A.22}$$

Putting together these two expressions, the inequality in (A.31) becomes:

$$\frac{\partial^2 P_{ijl}}{\partial p_{jl} \partial p_{kl}} - \frac{1}{P_{ijl}} \frac{\partial P_{ijl}}{\partial p_{jl}} \frac{\partial s_{ijl}}{\partial p_{kl}} =$$

$$-\frac{\alpha q_i^{1-\theta}}{1-\sigma} P_{ijl} P_{ij|l} P_{ik|l} ((1-\sigma)(1-P_{il}) - q_i^{-\theta}) - \alpha q_i P_{ijl} P_{ik|l} (1-\sigma)(1-P_{il}) \tag{A.23}$$

Equation (A.37) is positive, and hence a price equilibrium exists and is unique, if and only if, for all firms j, locations l, and customer types i:

$$P_{ij|l} \leq \frac{q_i^{\theta}(1-\sigma)(1-P_{il})}{(1-\sigma)(1-P_{il}) - q_i^{-\theta}} \tag{A.24}$$

(ii) Change in price by firms in a different location

For an increase in the price charged by firms operating in different locations, inequality (A.31) becomes:

$$\begin{aligned} -\frac{(\alpha q_i)^2}{1-\sigma} P_{ijl} P_{ikh} \Big(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta} - P_{il}) \Big) \\ + \frac{(\alpha q_i)^2}{1-\sigma} P_{ijl} P_{ikh} \Big(q_i^{-\theta} + P_{ij|l}((1-\sigma)(1-P_{il}) - q_i^{-\theta}) \Big) \geq 0 \\ \iff \frac{(\alpha q_i)^2}{1-\sigma} P_{ijl} P_{ikh} P_{il} \geq 0 \end{aligned} \tag{A.25}$$

which is always true, as the term on the left-hand side is always ≥ 0.

A.4 Marginal effect of number of firms on optimal prices

Without loss of generality, I show how prices change in response to the number of firms operating in a location when firms face a single type of consumers. Optimal prices under this assumption are simply:

$$p_{jl}^* = c_l + \left(\frac{1-\sigma}{\alpha [q_i^{1-\theta} + P_{ij|l}(q_i(1-\sigma)(1-P_{il}) - q_i^{1-\theta})]} \right) \tag{A.26}$$

Taking the total derivative of equation (A.23) with respect to the number of type-k firms operating in the the location:

$$\frac{\partial p_{jl}^*}{\partial N_{kl}} = \frac{\partial c_l}{\partial N_{kl}} + \frac{d}{dN_{kl}} \left(\frac{1-\sigma}{\alpha [q_i^{1-\theta} + P_{ij|l}(q_i(1-\sigma)(1-P_{il}) - q_i^{1-\theta})]} \right) \tag{A.27}$$

Let $\tilde{S} = q_i^{1-\theta} + P_{ij|l}(q_i(1-\sigma)(1-P_{il}) - q_i^{1-\theta})$. The derivative of this term with respect to the number of firms in the same location is:

$$\frac{d\tilde{S}}{dN_{kl}} = \left(\frac{\partial P_{ij|l}}{\partial p_{jl}^*}\frac{\partial p_{jl}^*}{\partial N_{kl}} + \frac{\partial P_{ij|l}}{\partial N_{kl}}\right)(q_i(1-\sigma)(1-P_{il}) - q_i^{1-\theta}) - \left(\frac{\partial P_{il}}{\partial p_{jl}^*}\frac{\partial p_{jl}^*}{\partial N_{kl}} + \frac{\partial P_{il}}{\partial N_{kl}}\right)P_{ij|l}q_i(1-\sigma) \tag{A.28}$$

Plugging this expression into equation (A.24) and rearranging:

$$\begin{aligned} &\frac{\partial p_{jl}^*}{\partial N_{kl}}\left[1 + \frac{1-\sigma}{\alpha\tilde{S}^2}\left(\frac{\partial P_{ij|l}}{\partial p_{jl}^*}(q_i(1-\sigma)(1-P_{il}) - q_i^{1-\theta}) - \frac{\partial P_{il}}{\partial p_{jl}^*}P_{ij|l}q_i(1-\sigma)\right)\right] \\ &= \frac{\partial c_l}{\partial N_{kl}} - \frac{1-\sigma}{\alpha\tilde{S}^2}\left(\frac{\partial P_{ij|l}}{\partial N_{kl}}(q_i(1-\sigma)(1-P_{il}) - q_i^{1-\theta}) - \frac{\partial P_{il}}{\partial N_{kl}}P_{ij|l}q_i(1-\sigma)\right) \end{aligned} \tag{A.29}$$

I focus on the right-hand side of the equation. Due to the presence of production externalities, the first term, $\frac{\partial c_l}{\partial N_{kl}}$ is negative. The sign of the second term (as well as of the term on the left-hand side of the equation) depends on the sign of the term in parenthesis, which can be rewritten as:

$$-\frac{\partial P_{ijl}}{\partial N_{kl}} - P_{ij|l}P_{ik|l}\left(1 - \frac{1}{q_i^{\theta}(1-\sigma)}\right) \tag{A.30}$$

Notice that the sign of this term is more likely to be positive, and hence prices are more likely to increase with the number of firms: (i) the stronger the size of the market-size effect relative to the market share effect, which determine the sign of the derivative $\frac{\partial P_{ijl}}{\partial N_{kl}}$; (ii) the larger the quantity purchased by the consumer q_i; (iii) the lower the similarity of products within a location, σ. Notice that (ii) and (iii) also contribute to making $\frac{\partial P_{ijl}}{\partial N_{kl}} > 0$, as shown by equation (9).

B Appendix B: Additional Data and Empirics

B.1 Location definition

The choice of the parish as the location unit takes into consideration the geographical dispersion of firms, which is important for customer search and outsourcing, and the political administration to which firms are subject, which can affect firms' production cost.

On the demand side, the model assumes that customers who visit a location are able to observe the characteristics of all the products sold by firms in that location, but cannot observe products sold by firms in neighbouring locations. Thus, to define the borders of a location, it is important to consider how far customers are willing to travel to search for products. In this setting, it is reasonable to assume that, once in a location, individuals in this setting walk around to search for products. Firms that belong to the same location must therefore be of walking distance to one another. The average size of Kampala parishes is 2.03 square-kilometer, which is a reasonable area to walk.

Although a pure spatial algorithm could be used to generate "search" clusters, it is important to also take into consideration the supply-side. In the model, firms within the same location have the same production cost, as they share the same TFP, and face the same outsourcing cost. In Uganda, parishes are under the administration of a chief who is responsible for tax collection, the implementation of national and local government policies and, in some instance, the settling of land disputes. All these factors are likely to affect the productivity in a given location. Considering the trade-off between geographical dispersion and production amenities, the identification of parishes as locations seems to be a reasonable choice.

B.2 Imputation of customers' location

Before describing the imputation procedure, it is important to understand what are the correlates of missing location. Table A11 shows the results from a regression of a dummy for missing location on a number of transaction and firm characteristics, controlling for parish (Column 1) and firm fixed effects (Column 2). I consider both transactions for which no information about the location is provided, and transaction for which only the region was recorded as missing.

Reassuringly, Column 1 shows that attrition is uncorrelated with the total amount of the transaction, the number of items purchased, and the firm's average number of daily customers and average daily revenues, suggesting that busier firms are not more likely to omit customers' origin. However, the table shows that location is more likely to be missing for final than for retail customers. A possible explanation is that retail customers are 21% more likely to have had interactions with the firm in the past, increasing the chances that the firm owner is aware of the origin of the customer. In Column 2, I include firm fixed effects to look at variation in reporting within the firm. Once again, I find no significant relationship between size of the transaction and attrition, but find that firms are less likely to report the origin of retail customers. Overall, these results show that, conditional on customer type, transactions for which customer origin is observed are not different from those for which location is missing in terms of observable characteristics. This suggests that they can be safely used to impute missing locations.

I rely on the structure of the model for imputation. Equation (7) shows that, within a location, the share of type q consumers buying products from a given firm is independent of the origin i of the customer. This is because mean utility δ^q_{jl} is independent from the customer's origin. This implies that, conditional on customer type and firm location, the distribution of customers' origin should not differ across transactions for which location is reported, and those for which it is missing in terms. Therefore, I randomly assign customers to locations proportionally to the share of customers origin observed in the data, conditional on customer type and firm location.

B.3 Additional Data for Estimation

B.3.1 Data for outside option

The estimation of retailers' outside option requires data on the total number of final and retail customers purchasing tailoring products in each parish. For final customers, I constructed this measure using data from the 2020 Ugandan National Panel Survey, which contains information on households' annual consumption of clothing. This information was used to calculate the share of households in Kampala purchasing tailoring products over a three days period (the length of the transaction data), assuming consumption is uniformly distributed over the year. The corresponding number at the parish level was then calculated by multiplying this share by the number of households per parish from 2014 Ugandan Population Census.

For retail customers, total number of customers was constructed combining data from the 2010 Ugandan Census of Business Establishments and the customer survey. Data from the latter shows that on average retail customers purchase products from a firm every 3.5 days. I therefore considered the total number of retailers as the pool of customers. I considered retail customers all firms operating in one of the following sectors: wholesale of textiles, clothing and footwear (ISIC 4641), retail sale of clothing, footwear and leather articles in specialized stores (ISIC 4771) and retail sale via stalls and markets of textiles, clothing and footwear (ISIC 4782). I used the 2010 Census, which includes geo-localized data, to compute the number of firms in these sectors operating in each parish.

B.3.2 Information about out of sample parishes

The number of firms in the parishes outside the sample is imputed using the 2010 Census of business establishments and assuming an entry rate equal to the average growth rate observed in the peripheral parishes in my sample. The number of owners from each parish is computed as the sum of all firms operating in a parish and the owners in our sample who are from the same parish, but operate in the core. Factor prices and productivity in peripheral parishes outside our sample are set equal to the average of the peripheral parishes in sample.

B.4 Mystery shoppers script

"Hi, I am looking for someone who can sew for me a short dress for my niece who is 13 years old girl. I got your contact from a friend who recommended you, and so I would like you to make the dress. Specifically, I would like you to reproduce this dress."

- Show the picture of the garment to be replicated to the tailor (Figure A12).

"As you can see, the dress has a gathered skirt, a baby collar and puff sleeves finished with elastics. On the back, it should have a long zip, not buttons. I bought some fabric that I would like you to use for this dress."

- Show the fabric to be used to the garment.

"Would you be able to do it? These are the measurements for the dress."

- Show the measurements to the respondent. Do not leave them with him/her.

"I am going to travel to Soroti in 3 days, and I would need the dress by then, so I can bring it with me. I will leave at [time at which you placed the order]. Would you be able to make it by then?"

If not: *"Why not? When would be the earliest you can make it?"*

- Accept the time frame given by the respondent as long as it is within the next 2 weeks.

"At what price would you be willing to sew this dress for me?"

- Reduce the price by 20%. If the reduced price is above 30,000UgSh, say that 30,000UgSh is the maximum you can offer.

"Would you be willing to sew it for me for [rounded price]?"

- Accept whatever price is then given by the respondent, as long as it is below or equal to 30,000UgSh. If not, thank the respondent and leave without buying anything.

"How much should I give you as a deposit?"

- Agree to deposit up to 50% of the price if the respondent insists.

"Can I please have a receipt, so that I can remember how much the balance is? Ok, then I will come and collect it on [earliest day available]. If you happen to finish the dress before, please give me a call at this number [phone number]"

- Give your phone number to the respondent. Leave the fabric and thank him/her. End of the exercise.

B.5 Additional drivers of agglomeration

In this section, I discuss additional potential drivers of firm agglomeration based on the evidence in Panel A of Figure 2.

Price collusion: An additional incentive for firms to operate near one another could be their ability to collude on prices. Two pieces of evidence suggest that this is not one of the key drivers of firm agglomeration in this context. First, only 7.5% of firms (8% and 4.6% in the core and periphery respectively) report benefiting from operating in proximity to other garment firms due to formal price agreements. Second, prices from the mystery shoppers exercise are, if anything, more dispersed in the core than in the periphery (Figure A4), which is inconsistent with collusion being more likely to occur in high-density areas.

Bargaining for better prices: Evidence from the mystery shoppers exercise indicates that customers are not able to bargain for better prices in the core. In the exercise, after asking for a starting price, interviewers were instructed to ask firms for a 20% discount (see Appendix B.4). I collected data on both the starting price and the price customers obtain after asking for a discount. On average, interviewers obtained a 12% discount on the initial price. However, I do not find a significant difference in the discount given by firms in the core and the periphery, both in absolute value and as a percentage of the initial price.

Proximity to other sector firms: Only a few firms across the core and periphery, 14% and 7% respectively, mention proximity to other-sector firms as affecting their location choice. These figures are considerably lower than the percentage of firms citing proximity to other garment businesses (28% and 11% respectively), confirming that the most relevant dimension of agglomeration is within the sector. However, I explicitly incorporate proximity to other sector businesses as a possible driver of agglomeration in the model.

Affordable rent: In Figure 2, affordable rent appears to be an equally important driver of location choice for businesses in the core and in the periphery. This may be surprising given the significantly higher rental prices paid by firms in the core compared to the periphery. However, firms in the core manage to mitigate these costs by renting much smaller premises, often a space within a room with other garment firms. As a result, total rental costs are similar across the two locations. The ability of firms in the core to share premises is implicitly included in the model as one of the forces driving lower marginal costs in this location.

Gender preferences for commuting: The data does not show evidence of a gender gap in commuting (Le Barbanchon et al., 2021): the percentage of female firm owners is 75% in the core and 74% in the periphery. Figure A5 replicates Figure 2 separately for men and women. It shows that the determinants of firms' choice of location are similar across genders, both in absolute value and in terms of differences between core and periphery.

B.6 Outsourcing

In this section, I provide evidence on a feature of firms' production process that is suggestive of the presence of production externalities in this setting. Data from the firm survey indicates that outsourcing is very common in the Ugandan garment sector, with 40% of the production carried out by external workers. The most commonly outsourced tasks are overlocking (50% of firms), making buttonholes (20%), and ironing (13%).

Panel A of Table A12 shows summary statistics on outsourcing, separately for firms located in the core and the periphery. On average, firms in the core: (i) employ more workers in production; (ii) are 14 percentage points (24%) more likely to employ at least one external worker; (iii) outsource a larger share of their production to external workers (42% vs. 32% in the periphery). The fact that outsourcing is more common in the core could explain why these firms employ fewer workers, own fewer assets and operate on smaller premises than firms in the periphery (Table 1).

Panel B of Table A12 suggests that differences in outsourcing across core and periphery arise because suppliers of intermediate tasks are also geographically concentrated in the core. For instance, more than 95% of task providers in the core are located within a 5-minute walking distance from the firm, compared to only 56% in the periphery . Proximity to suppliers reduces both transport costs and information frictions by minimizing the opportunity cost of transporting and retrieving products from suppliers, and by facilitating monitoring. Instead, there is no evidence that firms in the core are more likely to be operating at capacity. When asked by how much firms could increase their production without hiring additional workers or purchasing additional machines, 6% of businesses across the core and the periphery say they could double it or more, and 31% say they could increase production slightly. Only 8% of firms in the core and 12% of firms in the periphery report they could not increase their production.

These types of firm-to-firm interactions are comparable to the machines rental market studied by Bassi et al. (2022b) in the Ugandan carpentry sector, which provide firms with a workaround for the market imperfections that prevent investments in machinery. Similar constraints are likely to apply to the garment sector. However, machine rental is not very common in this sector: only 8% of businesses in the periphery and 10% in the core rent any machines. A possible explanation for the difference in firm behavior across the two sectors is the different size of machines used in production. The average value of a machine in the carpentry sector is $1,642, compared to only $125 for a sewing machine and $222 for an overlocking machine, which are the most common assets utilized in the garment sector. The primary constraints in garment are likely to be the cost of renting space for additional machines or workers, and the indivisibility of labor. In fact, the most commonly outsourced tasks typically take less time than the sewing. Given the small size of firms, this means that an additional worker (for whom the firm would have to rent larger premises) would experience significant idle time unless shared by several businesses.

C Appendix C: Robustness and Extensions

C.1 Allowing for dynamics

To model presented in Section 4 is static. Although this is partly justified by the persistence of firm-customer relationships and firm location choices in the data, I conduct a robustness check to test the plausibility of this assumption by adding a second period to the model. In the first period, consumers do not observe any of the match-specific shocks ε and decide what firms to buy products from as described in Section 4.2. However, upon visiting a location, consumers observe the ε-shocks for the firms they visited in the selected location. Hence, in a second period, consumers would have to decide whether to go back to the same location (in which case they would buy from the same firm, as the search strategy and reservation values would be unchanged), or visit a different location.

Let $l^{(1)}$ and $j^{(1)}$ be the location and the firm chosen in the first period. In the second period, consumers will choose the location that yields the highest expected utility:

$$l^{(2)} = \arg\max_{k \in L}\{u_{ij^{(1)}l^{(1)}}, \max_{k \neq l^{(1)}} V_{ik}\} \tag{A.31}$$

where V_{ik} is given by equation (8). I simulate a second period using the estimated parameter and compute the share of customers who would choose to search a different location. I find that only 10% of customers would switch to a different location in a subsequent period, while 90% would go back to the initial firm. This suggests that a static model captures the core of the consumer search process and hence of firm choice of location.

C.2 Model extension with entry

C.2.1 Set-up

In this section I present an extension to the model that allows for firm entry and exit. Let E be the number of potential entrants, which is finite and known to all firms, and let J be the corresponding number of firms actually entering the market. Potential entrants simultaneously choose whether to enter the market and in which location to place the firm. I assume that firms observe their type (high or low-quality) prior to making the entry decision. To enter, firms must pay an entry cost EC_j. If they decide not to enter the market, they make zero profits. Firms' total profits are given by the following expression:

$$\Pi_{jl}(\mathbf{L},\mathbf{p}) = \pi_{jl}(\mathbf{L},\mathbf{p}) - \tau_2 g_2(||z_j - z_l||) - e_{jl} - EC_j \tag{A.32}$$

In equilibrium, each entrant expects to earn non-negative profits. Given the assumption on the unobserved preference shock e_{jl}, the probability Pr_j of a firm entering the market is:

$$Pr_j(\mathbf{L}, \mathbf{p}) = \frac{\sum_{l=1}^{L} \exp\left((\pi_{jl}(\mathbf{L}, \mathbf{p}) - \tau_2 g_2(||z_j - z_l||) - EC_j)\frac{1}{\lambda} \right)}{1 + \sum_{l=1}^{L} \exp\left((\pi_{jl}(\mathbf{L}, \mathbf{p}) - \tau_2 g_2(||z_j - z_l||) - EC_j)\frac{1}{\lambda} \right)} \tag{A.33}$$

where the term at the numerator is a weighted average of firms' expected profits across all locations. Notice that the expected number of entrants depends on firms' location in equilibrium, which in turn depends on the expected number of entrants. I solve for equilibrium entry and location by searching for a fixed point to the augmented system of equations (17) and (A.42) for all potential entrants.

C.2.2 Estimation of entry costs

Under the assumption that the expected number of entrants is exactly equal to the number of entrants in the data, it is possible to recover firms' entry costs following the approach outlined in Seim (2006). In the model, firms are only heterogeneous in quality and the location where the owner resides. Let E_j be the potential number of type-j entrants, and let $J_j = Pr_j(\mathbf{L}, \mathbf{p}) \times E_j$ be the actual number of type-j entrants. To recover entry costs, one must know the number of potential entrants from each location. I assume that this number is equal to 2% of the population and that there is an equal share of low and high-quality type owners in each location (i.e. each parish). Notice that entry costs can be expressed as:

$$-\frac{EC_j}{\lambda} = \log Pr_j(\mathbf{L}, \mathbf{p}) - \log(1 - Pr_j(\mathbf{L}, \mathbf{p})) - \log \sum_{l=1}^{L} \exp\left((\pi_{jl}(\mathbf{L}, \mathbf{p}) - \tau_2 g_2(||z_j - z_l||))\frac{1}{\lambda} \right) \tag{A.34}$$

The logarithm of the probability of entry can be re-written as $\log Pr_j(\mathbf{L}, \mathbf{p}) = \log J_j - \log E_j$, and the probability of *not* entering as $1 - Pr_j(\mathbf{L}, \mathbf{p}) = \log(E_j - J_j) - \log E_j$. Plugging these expressions into (A.43), I obtain:

$$-EC_{ot} = \lambda \times \left[\log J_j - \log(E_j - J_j) - \log \sum_{l=1}^{L} \exp \left((\pi_{jl}(\mathbf{L}, \mathbf{p}) - \tau_2 g_2(||z_j - z_l||)) \frac{1}{\lambda} \right) \right] \tag{A.35}$$

Given the estimated parameters, the number of potential and actual entrants and equilibrium $\mathbf{p}$ and $\mathbf{L}$, entry costs can be directly calculated from (A.44). I estimate average entry costs for low-quality and high-quality firms to -120.2 and -123.7 respectively. These costs are very close to the average price of a sewing machine in the data ($132.5), which is the only asset firms need to purchase to start a business.

Appendix Tables

Table A1: Attrition

	(1) Filled transaction record	(2) Listed any customer
Core	-0.034	0.072
	(0.028)	(0.051)
Number of employees	0.004	-0.003
	(0.019)	(0.030)
Log(Monthly revenues (USD))	-0.025	-0.017
	(0.018)	(0.031)
Number of weekly customers	-0.002	0.005**
	(0.002)	(0.002)
Mean outcome	0.908	0.639
Number of Observations	601	601

Notes: Standard errors clustered at the parish level are reported in parenthesis. * $p < .10$, ** $p < .05$, *** $p < .01$. Data is from transaction records (Columns 1 and 2) and the list of customers recorded by enumerators during the firm interview (Column 2). The outcome in Column 1 is a dummy equal to one if the firm owner filled the transaction records. The outcome in Column 2 is a dummy equal to one if either a customer was listed during the interview, or the firm recorded the contact details of at least one customer in the transaction records.

Table A2: Relocation

	% of firms
No Relocation	
Never relocated	54.4
Relocation to Core	
Periphery to core	5.32
Outside Kampala to core	6.16
Relocated within core	11.3
Relocation to Periphery	
Core to periphery	2.83
Outside Kampala to periphery	7.82
Relocated within periphery	12.1

Notes: Table A2 shows the percentage of firms that never relocated, relocated to the Core and relocated to the Periphery since being established. Data is from the firm baseline survey.

Table A3: Consumers' search methods

	% of final consumers	% of retailers
Walk into any firm	53.5	55.8
Ask friends/family members	43.9	42.4
Ask other garment firm	38.6	51.1
Ask firm in different sector	6.9	11.9
Look on the internet	7.9	4.0

Note: Data is from the customer survey. Table A3 shows the percentage of final customers and retailers who report they would be using a given method in the hypothetical scenario in which their current supplier shut down and they had to search for a new firm.

Table A4: Number of visited firms prior to purchasing

	(1) Number of visited firms	(2) Number of visited firms	(3) Number of visited firms
Core	0.570***	0.578**	0.696*
	(0.198)	(0.274)	(0.362)
Quality score		-0.041	-0.057
		(0.132)	(0.135)
Mystery shoppers price		0.059	0.047
		(0.076)	(0.081)
Product type FEs		✓	✓
Customer type FEs			✓
Mean in Periphery	1.821	1.821	1.821
N. of observations	556	491	491

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. Robust standard errors in parentheses. Table A4 shows the results from a regression of the number of firms visited by the customer prior to purchasing on a dummy equal to one for firms in the core. The unit of observation is the firm-customer match.

Table A5: Learning

	(1) Quality score	(2) Price of 1 m of fabric	(3) Delivered on time
Core	0.276*	0.821***	-0.017
	(0.162)	(0.254)	(0.054)
Years of experience	0.013*	0.010	0.000
	(0.007)	(0.013)	(0.003)
Years of experience X Core	-0.006	-0.030	-0.002
	(0.012)	(0.023)	(0.004)
Interviewer FEs	✓	✓	✓
Mean in periphery	-0.078	2.871	0.807
Number of observations	494	476	494

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. Robust standard errors in parentheses. Table A5 shows the results from a regression of the three quality measures on a dummy equal to one for firms in the core. In Column 1 the outcome is the quality score from the mystery shoppers exercise. The score is a standardized measure with mean 0 and variance 1. In Column 2 the outcome is the price of one meter of the typical fabric utilizes to produce garments in the month prior to the survey. In Column 3 the outcome is a dummy equal to one if the firm delivered the product by the date agreed with the mystery shopper.

Table A6: Estimation of price coefficient, first stage

	(1) Transaction price	(2) Delta
Panel A: First Stage		
Cost of zipper, final consumers	5.053***	
	(0.854)	
Cost of zipper, retailers	8.298***	
	(2.962)	
Panel B: Second Stage		
Transaction price, final consumers		-0.107***
		(0.021)
Transaction price, retailers		-0.073**
		(0.029)
Product FEs	✓	✓
Number of Observations	2,367	2,367
Weak identification test: Cragg-Donald Wald F statistic = 15.398		

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. Robust standard errors in parentheses. Table A6 reports the results from a two-stage least squares regression of the mean utility δ on instrumented prices from transaction records. The instrumental variables for prices are the average cost paid by the firm for a zipper.

Table A7: Estimated demand parameters using convex hulls

	Parameter	Administrative Boundaries	Convex Hull
PANEL A: Demand			
Price final consumers	α_f	-0.107	-0.105
Price retailers	α_r	-0.071	-0.070
Quality final consumers	β_f	0.284	0.278
Quality retailers	β_r	0.289	0.281
Taste shocks correlation	σ	0.578	0.585
Quantity multiplier	θ	0.379	0.388
Travel cost	τ_1	1.399	1.416
Within location search cost	μ	0.0016	0.0010
Other-sector firms FE	γ_l	0.361	0.466

Note: Table A7 shows point estimates for parameter estimates of the model where the area of a parish is defined as (i) its administrative boundaries (as in the baseline estimation) or (ii) the convex hull of the firms located in that parish.

Table A8: Robustness of τ_2 estimates to different values of λ

λ	τ_2	Share high-quality in core	Share low-quality in core
$0.75 \times \bar{\pi}$	-7.799	0.501	0.412
$1.00 \times \bar{\pi}$	-8.009	0.496	0.432
$1.25 \times \bar{\pi}$	-8.257	0.492	0.443

Notes: Table A8 reports the estimated commuting parameter τ_2 for different values of λ. $\bar{\pi}$ are average variable profits across all firms and locations. In the last two columns, the table reports the average share of high-quality and low-quality firms locating in core at the corresponding equilibrium. The average share of high and low-quality firms in the core observed in the data are 0.504 and 0.448.

Table A9: Model fit—Observed and simulated inputs

Parish		Land (h)	Labor (ℓ)
Bwaise II	*Data*	*6.050*	*1.931*
	Sim	1.548	1.373
Kamwokya II	*Data*	*5.450*	*1.650*
	Sim	3.860	1.718
Kasubi	*Data*	*5.003*	*2.246*
	Sim	4.955	3.439
Katwe I	*Data*	*1.750*	*1.500*
	Sim	1.524	8.195
Kibuye II	*Data*	*2.857*	*2.429*
	Sim	1.473	1.499
Kisenyi III	*Data*	*3.450*	*2.450*
	Sim	1.876	4.424
Kisenyi II	*Data*	*2.783*	*2.256*
	Sim	2.087	1.766
Kisenyi I	*Data*	*2.366*	*1.781*
	Sim	3.779	3.197
Kisugu	*Data*	*7.750*	*1.938*
	Sim	7.654	2.045
Mbuya I	*Data*	*9.394*	*2.314*
	Sim	2.712	2.029
Naguru I	*Data*	*3.862*	*2.353*
	Sim	4.881	0.730
Nakasero IV	*Data*	*2.789*	*2.650*
	Sim	1.055	0.892
Nakivubo-Shauriyako	*Data*	*4.533*	*2.467*
	Sim	0.753	1.850
Wandegeya	*Data*	*2.478*	*2.696*
	Sim	1.106	1.428

Notes: Table A9 shows the average amount of land and labor employed by firms across different locations from the firm survey data (Data) and the simulated data (Sim). Land is measured as the size of the firm premises in square-meters. Labor is measured as the total number of workers employed by the firm, including the firm owner.

Table A10: Profits and welfare in the E-commerce counterfactual

	Baseline	**E-commerce**
Share of firms in core	0.466	0.315
		-32.3%
Share of high-quality in core	0.496	0.301
		-39.3%
Share of low-quality in core	0.432	0.331
		-23.5%
Share of sales in core	0.482	0.205
		-57.5%
Average price	9.002	9.20
		2.17%
Average profits	96.49	104.6
		8.44%
Average consumer welfare	45.94	59.90
		30.4%

Note: Table A10 shows firm location, prices, profits and consumer welfare at the baseline equilibrium and in the E-commerce counterfactual.

Table A11: Correlates of missing customer origin

	(1)	(2)
	Missing origin	Missing origin
Total transaction value (USD)	0.000	-0.000
	(0.001)	(0.000)
Quantity for customer	0.000	-0.000
	(0.000)	(0.000)
Retail customer	-0.071*	-0.069***
	(0.035)	(0.023)
Average of daily customers	-0.041	
	(0.039)	
Average daily revenues (USD)	-0.001	
	(0.001)	
Parish FEs	✓	
Firm FEs		✓
Number of Observations	2,569	2,569

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. Standard errors clustered at the parish level in parentheses. Table A11 reports the results from a regression of a dummy equal to 1 if the origin of the customer is missing in the transaction records on a number of firm characteristics. The regressions in columns 1 and 2 include parish fixed effects and firm fixed effects respectively.

Table A12: Outsourcing

	Core	Periphery	P-value
PANEL A: Outsourcing			
Total number of workers	2.240	1.927	[0.000]
Any external worker	0.726	0.583	[0.000]
Share of external workers	0.418	0.324	[0.000]
PANEL B: Distance from Suppliers			
Within 5 minutes	0.954	0.557	[0.000]
Between 5 and 15 minutes	0.040	0.188	[0.000]
More than 15 minutes	0.005	0.257	[0.000]

Note: Data is from the baseline survey of garment firms. Table A12 shows means and p-values from a t-test of equality of means across core and periphery, controlling for product type fixed effects. Panel A shows the total number of workers employed in the production of the most typical product sold by the firm, the share of firms that employ at least one external worker, and the total share of external workers employed in production. Panel B shows the average walking distance to the majority of external workers employed by the firm.

Table A13: Estimated parameters without information frictions

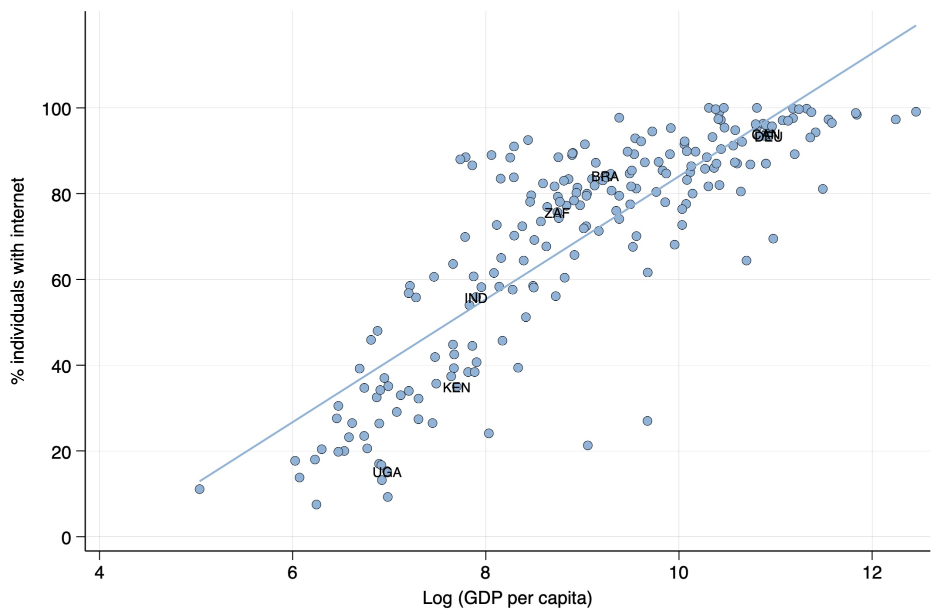
	Parameter	**Estimate**	**Std Error**
PANEL A: Demand			
Price final consumers	α_f	-0.083	(0.017)
Price retailers	α_r	-0.045	(0.020)
Quality final consumers	β_f	0.222	(0.050)
Quality retailers	β_r	0.196	(0.068)
Taste shocks correlation	σ	0.668	(0.016)
Quantity multiplier	θ	0.288	(0.008)
Travel cost	τ_1	0.890	(0.038)
Other-sector firms FE	γ_l	-0.015	(0.001)
PANEL B: Supply			
Labor share	δ	0.670	(0.018)
Baseline productivity	T_1	2.081	(0.910)
Production externalities	T_2	0.132	(0.268)
PANEL C: Location			
Commuting cost	τ_2	5.736	(1.195)

Note: Table A13 shows point estimates and standard errors for parameter estimates of a discrete choice demand model with no information frictions. Standard errors are bootstrapped using 100 bootstrapped samples.

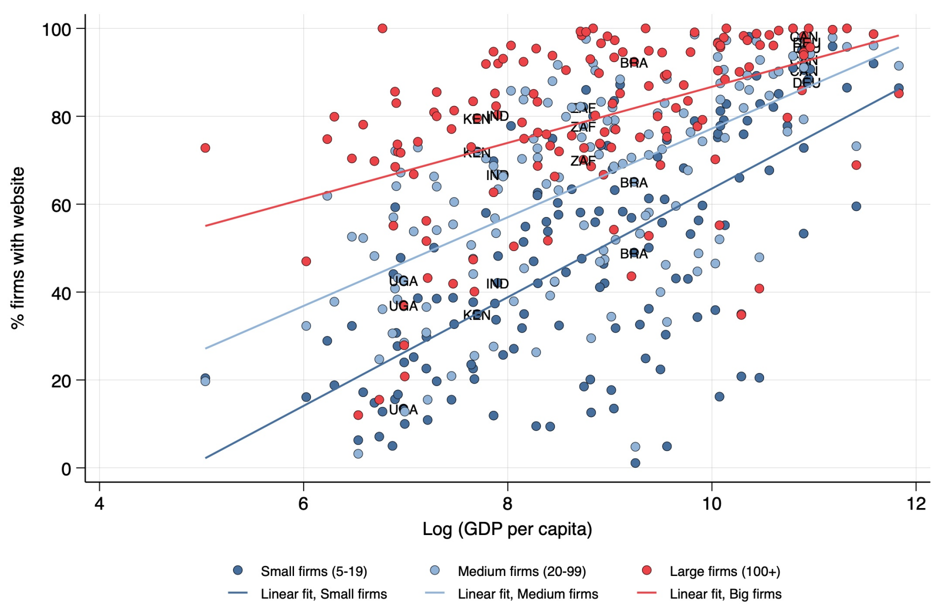
Appendix Figures

Figure A1: Internet usage and GDP per Capita

PANEL A: % of individuals using internet



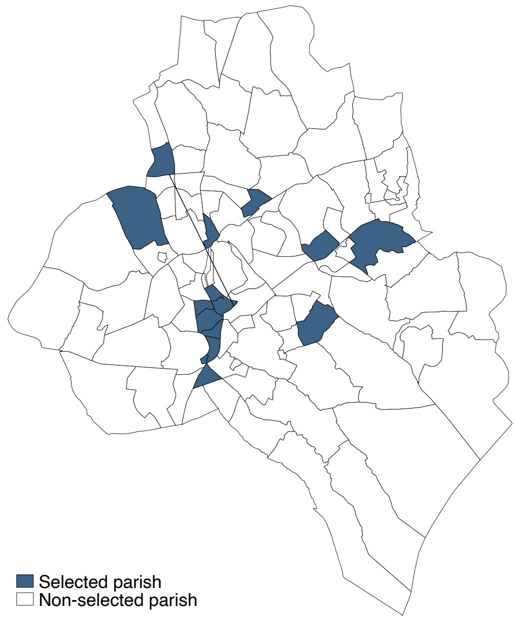
PANEL B: % of firms with own website



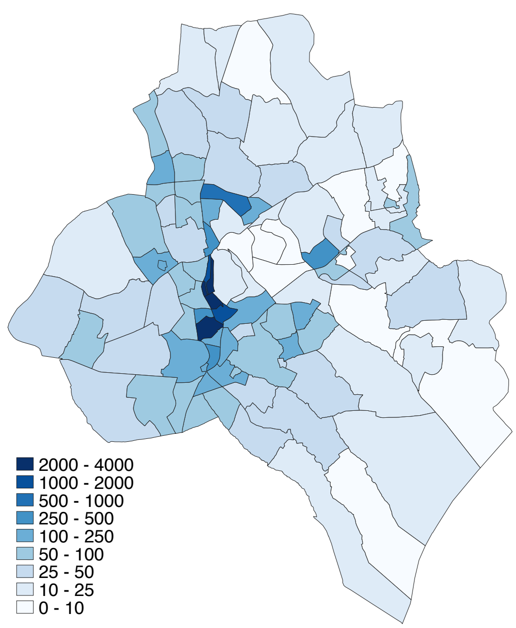
Note: Panel A of Figure A1 plots the percentage of individuals with internet access (source: World Telecommunication/ICT Indicators Database, International Telecommunication Union) against the logarithm of GDP per capita. Panel B shows the percentage of formal firms with their own website (source: World Bank Enterprise Surveys), also plotted against the logarithm of GDP per capita.

Figure A2: Selected locations

PANEL A: Selected Parishes

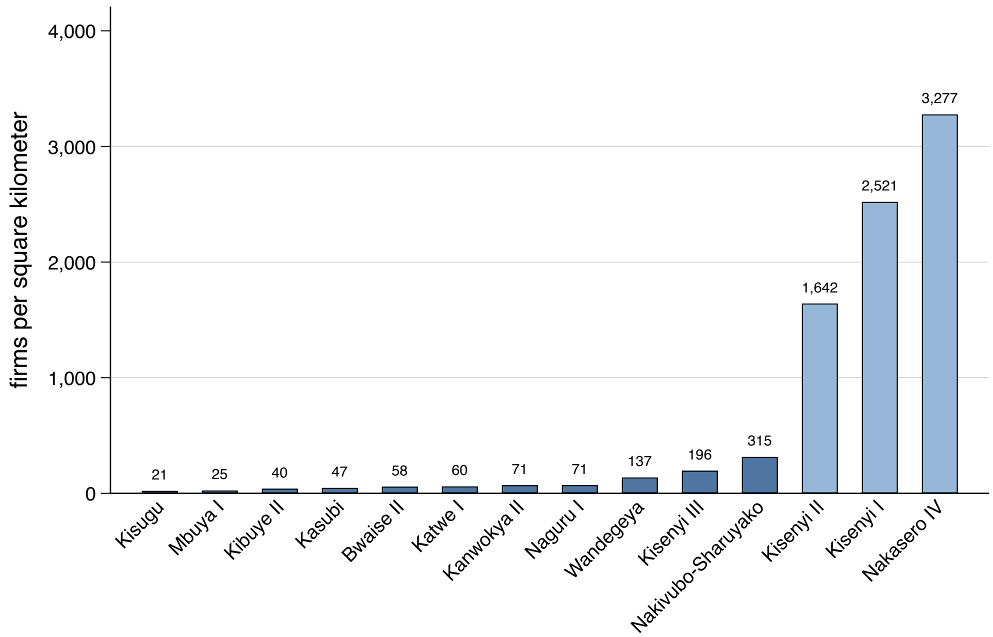


PANEL B: Firm density



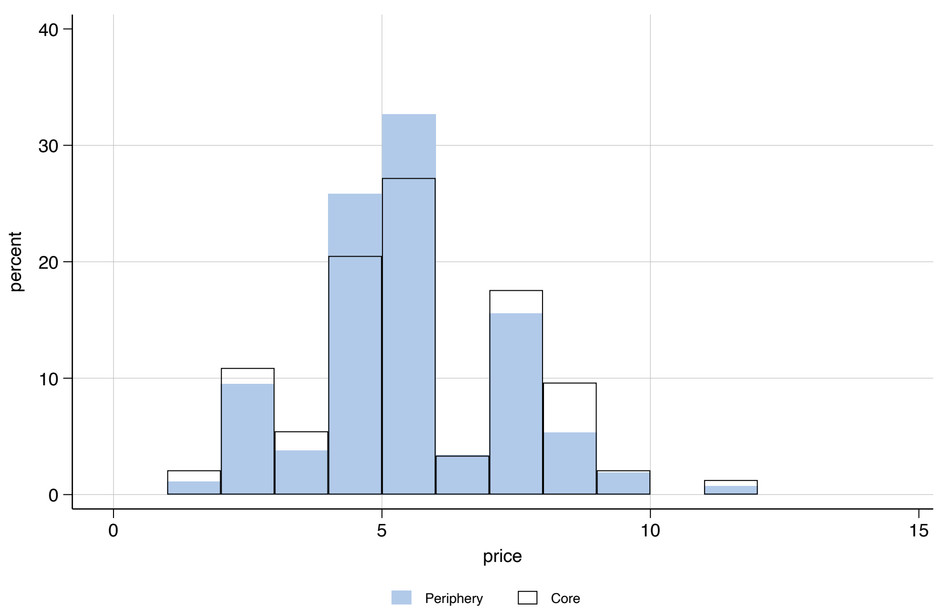
Note: Panel A of Figure A2 shows the 14 parishes selected for the study. Panel B shows the firm density, measured as the average number of firms per square kilometer, across the 96 parishes in Kampala.

Figure A3: Number of firms per square kilometer in selected parishes



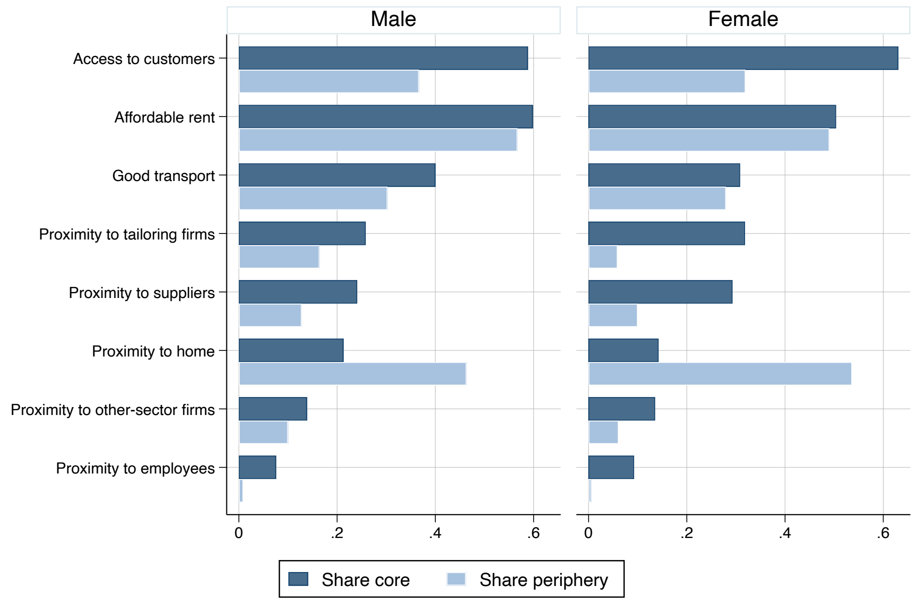
Notes: Figure A3 shows the number of garment firms per square kilometer in each of the selected parishes using data from the census of garment firms. Parishes in the periphery are in dark blue, while parishes in the core are in light blue.

Figure A4: Distribution of mystery shoppers prices



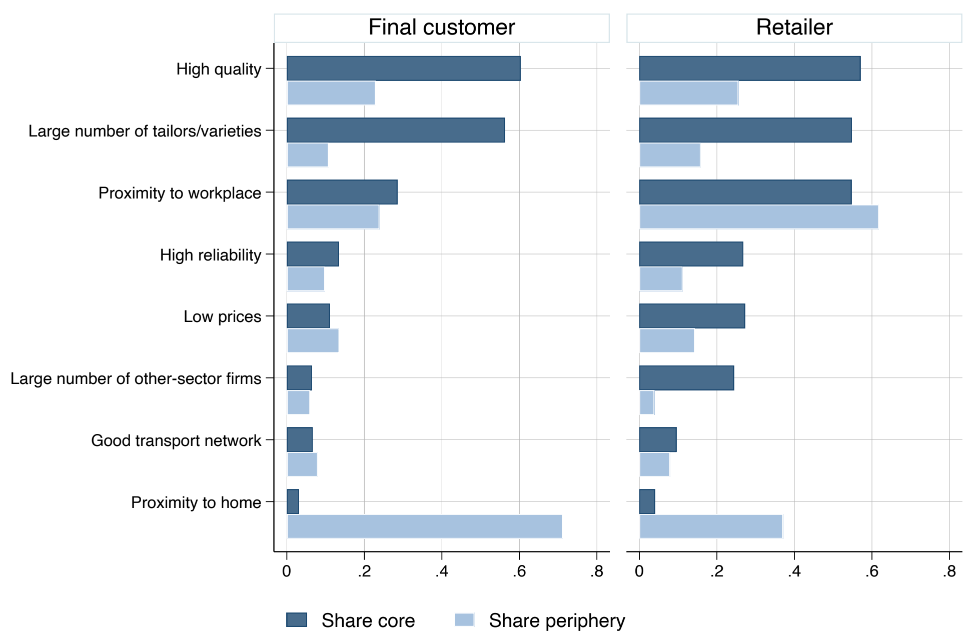
Note: Figure A4 shows the distributions of prices (in USD) from the mystery shoppers exercise separately for firms in the core (in black and white) and the periphery (in light-blue).

Figure A5: Reason for *locating* in Core vs. Periphery, by Gender



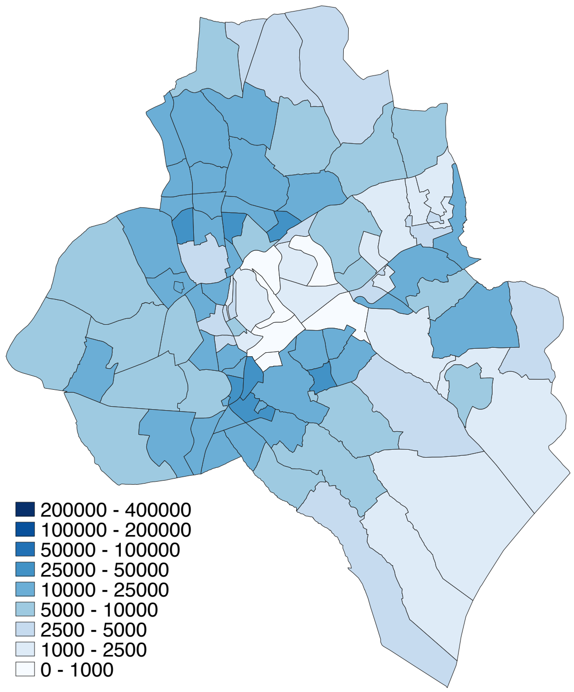
Note: Data is from the baseline survey of firms. The sample includes 154 male firm owners and 447 female firm owners. The dark blue bars show the share of male and female firm owners in the core reporting the reason indicated on the left as a driver of their initial choice of location. The light blue bars show the same statistic, but for firms operating in the periphery.

Figure A6: Reason for *searching* in core vs. periphery, by customer type



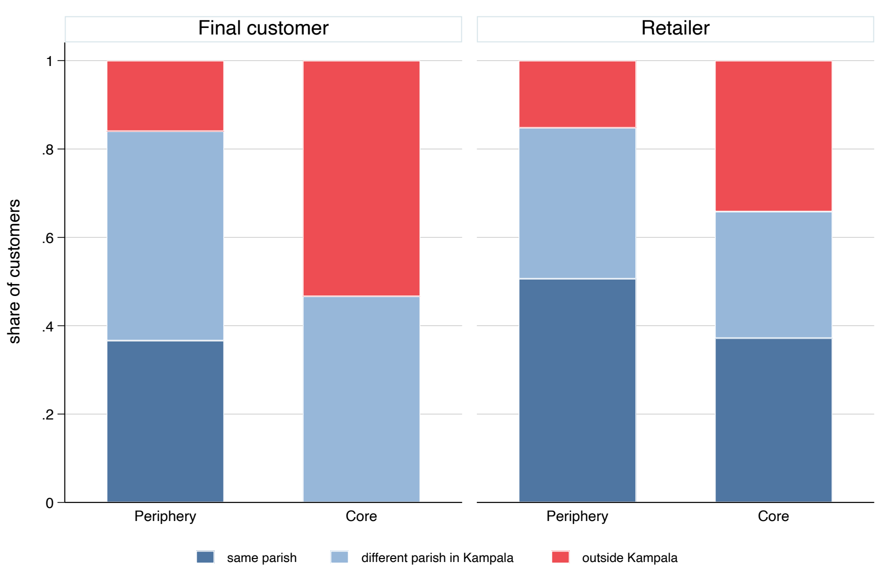
Note: Data is from the baseline survey of customers. The sample includes 303 final customers and 278 retailers. The dark blue bars show, among customers who indicated they would prefer to search for a new firm in the core, the share reporting the reason indicated on the left as a driver of their choice of location. The light blue bars show the same statistic, but for customers who indicated they would prefer to search for a firm in the periphery.

Figure A7: Population density



Note: Figure A7 shows the population density, measured as the number of inhabitants per square kilometer, across the 96 parishes in Kampala. Population density is computed using data from 2014 National Population and Housing Census. The relative intensity of the colors is the same as in Panel B of Figure A2 to facilitate the comparability of firm and population density.

Figure A8: Customer origin location, by Customer Type

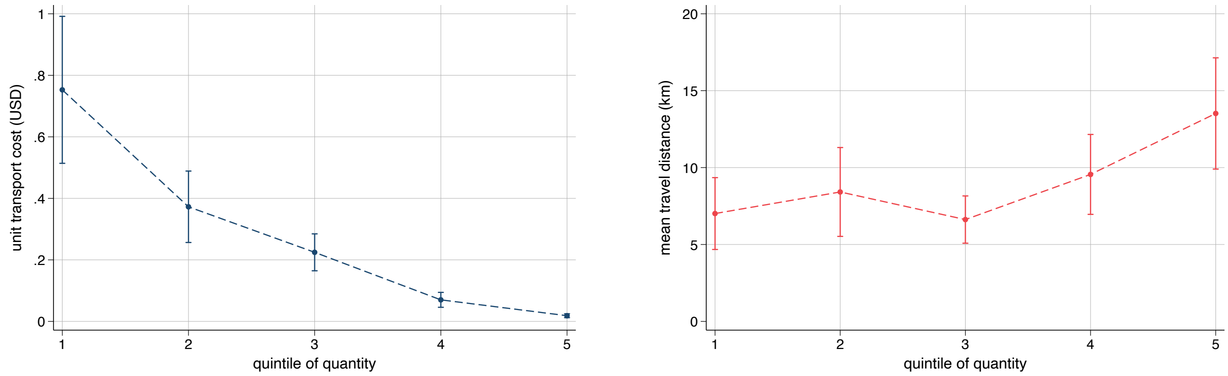


Note: Figure A8 shows the distribution of customers' origin location, separately for final customers and retailers. Data is from the customer survey. For final customers, the origin is the home parish. For retailers, the origin is the parish where the retail business is located.

Figure A9: Unit transport cost and travel time, by quantity quintile

PANEL A: Unit transport cost

PANEL B: Travel distance

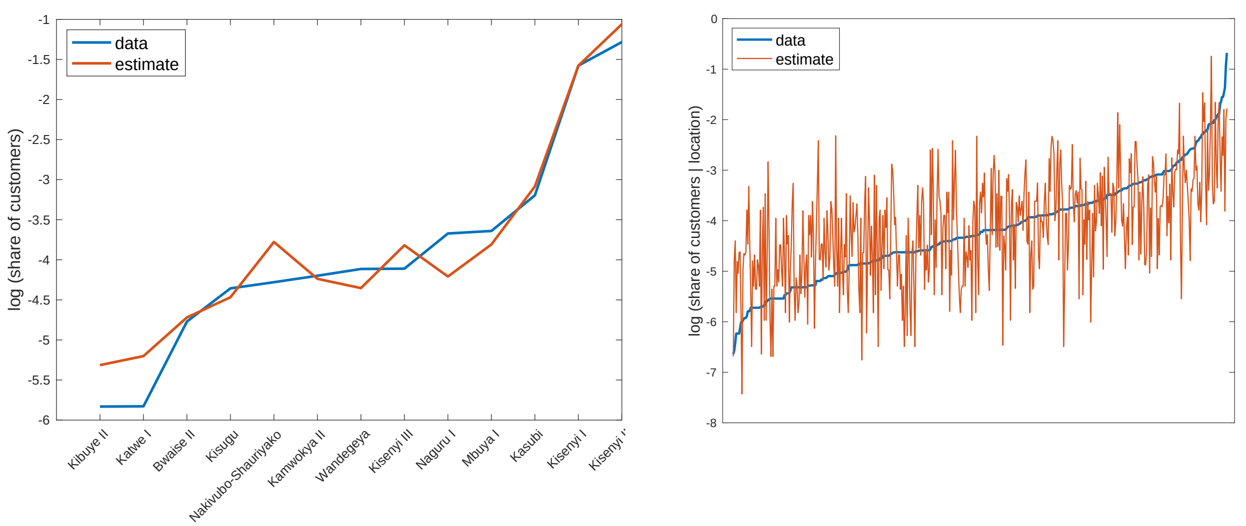


Notes: Panel A of Figure A9 plots the average unit transport costs (total transport costs divided by the quantity of goods purchased in a given transaction) and the corresponding 95% confidence intervals for each quintile of the distribution of purchases quantities. Panel B shows average travel distance in kilometers and correpsonding confidence intervals by quintile of the distribution of the purchased quantities. Data is from the survey of customers.

Figure A10: Model fit—Observed and simulated demand

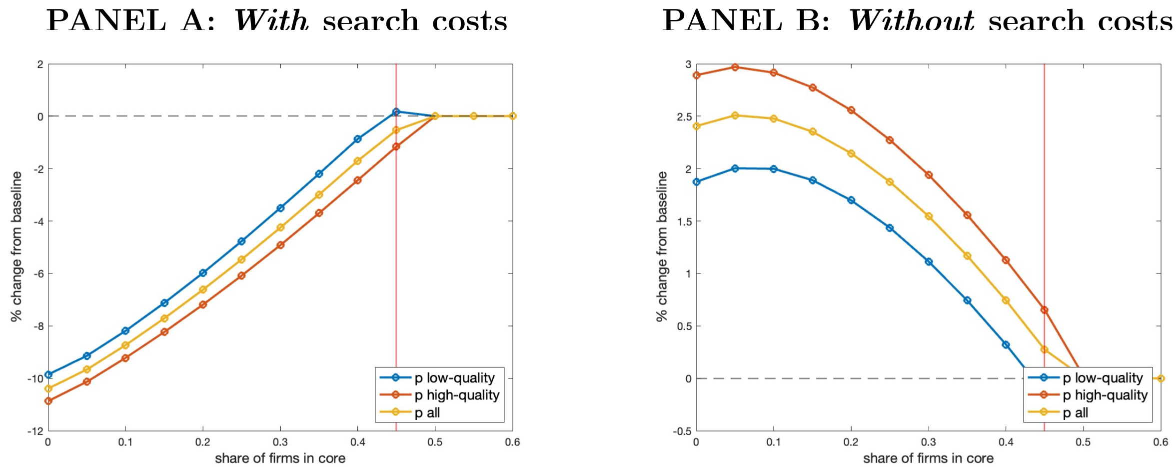
PANEL A: Location shares

PANEL B: Within location firm shares



Notes: Figure A10 shows the logarithm of location shares and within location firm shares from the firm transaction records (in blue) and the simulated data (in red).

Figure A11: Impact of evictions on prices



Notes: Figure A11 shows percentage changes in prices when only the share of firms indicated on the x-axis is allowed to operate in the core, relative to the baseline scenario with no restrictions. Panel A shows predicted changes from a model with within-location search costs. Panel B shows predicted changes from a model without such costs. In both panels, the blue and red lines report price changes among low- and high-quality firms respectively.

Figure A12: Product commissioned by mystery shoppers

Note: Figure A12 shows the product commissioned by the mystery shoppers to firms. This picture was shown to firms, who were asked to exactly replicate the dress. Photo credit: Mariajose Silva Vargas.

Figure A13: Quality scoring sheet

BUSINESS ID:				
	ASSESMENT CRITERIA	**SCORING GUIDE**	**MAX SCORE**	**SCORE**
1	DARTS	Dart of 4 ”long by 1”wide	3	
		Correctly sewn	3	
		Press to the right side	2	
		Position of the Dart observed	2	
2	COLLAR	Peter Pan/Baby Collar	5	
		Fixed correctly round the neckline	5	
3	SLEEVES	Sleeved Well Gathered	3	
		Sleeve Length 8”	2	
		Round sleeve 14”	2	
		Correctly fixed on Bodice	3	
4	SKIRT	Skirt length 18”	2	
		Skirt Equally Gathered	2	
		Neatly fixed to Bodice	2	
		Correct Seam Allowance	2	
		Skirt bottom shaped round	2	
5	ZIP	Zip attached to Centre back seam	4	
		Right color of Zip	3	
		Right length of Zip	3	
6	SEAM	Right Seam Allowance “Y2-1”	3	
		Correctly Pressed	3	
		Neatly Finished Edges	4	
7	HEM	Hemmed bottom of Dress	2	
		Hem lin-2ins	1	
		Hem Neatly sewn	3	
		Hem well pressed	4	
8	MEASUREMENTS	Cross Back 15”	2	
		Bust - 34”	2	
		Waist – 28”	2	
		Top to Waist –14”	2	
		Full Length – 32”	2	
9	FINISHING	No hanging threads seen	3	
		Dress Pressed with no wrinkles seen	3	
		No chalk marks	2	
		Dress clean	2	
10	PACKAGING	Dress Neatly and Correctly Folded	5	
		Packed in Bag	3	
		Branded Packaging	2	

Note: Figure A13 shows the assessment sheet used to evaluate the quality of the products purchased in the mystery shoppers exercise. Each product is evaluated according to 10 assessment criteria, which define the macro-categories for the assessment. Each criteria is then sub-divided in smaller categories which define the task that should have been accomplished by the firm. Each task is associated with a maximum score. The maximum scores add up to 100.