NBER WORKING PAPER SERIES

THE GEOGRAPHY OF CONSUMPTION

Sumit Agarwal
J. Bradford Jensen
Ferdinando Monte

Working Paper 23616
http://www.nber.org/papers/w23616

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2017, Revised March 2019

We would like to thank David Atkin, Victor Couture, Jonathan Dingel, Colin Hottman, Nate Miller, Steve Redding, Andrés Rodriguez-Clare, Esteban Rossi-Hansberg, Tony Yezer, and seminar participants at Columbia, Harvard-MIT, National University of Singapore, NBER, Philadelphia and Richmond FED, Singapore Management University, Southern Methodist University and University of Virginia for helpful comments and discussions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w23616.ack

The Geography of Consumption
Sumit Agarwal, J. Bradford Jensen, and Ferdinando Monte
NBER Working Paper No. 23616
July 2017, Revised March 2019
JEL No. F1,F14,L8,R1,R2

ABSTRACT

This paper examines the interaction between consumers' willingness to travel and producers' choices for a broad range of industries that supply final consumption and account for a large fraction of employment in the United States. Using detailed credit card data, we present evidence that consumers actively manage the spatial dimension of their purchases. Further, the data exhibit considerable variation in expenditure gravity across sectors. We develop a simple theory of how a sector characteristic, the durability/storability of the sector's output, affects consumer and producer behavior. We present empirical evidence that durability/storability appears to influence local employment, producer density, and establishment size differentially across sectors in U.S. counties. Our results have implications for a broad range of issues, including the consequences of local shocks, the impact of place-based policies, and the geographic market definition in the assessment of horizontal mergers.

Sumit Agarwal
National University of Singapore
Mochtar Raidy Building
15 Kent Ridge Drive
Singapore
ushakri@yahoo.com

Ferdinando Monte
Georgetown University
McDonough School of Business
37th & O Streets, NW
Washington, DC 20057
ferdinando.monte@georgetown.edu

J. Bradford Jensen
McDonough School of Business
Georgetown University
Washington, DC 20057
and Peterson Institute for International Economics
and also NBER
jbj24@georgetown.edu

1 Introduction

In spite of the importance of consumers' willingness to travel to the location and production decisions of many firms, surprisingly little is known about general features of consumer mobility, their variation across sectors, and their equilibrium impact on producers' choices. This paper documents considerable variation in gravity across sectors, identifies a sector characteristic that provides a plausible explanation for the variation, and presents evidence that local employment, producer density, and establishment size in the sector are likely influenced by this attribute.

Characterizing the geographic scope of local consumption markets has far-reaching implications for our understanding of a broad range of policy-relevant questions. Around 40% of employment in modern economies, from apparel stores to gasoline stations to restaurants, is involved in the delivery of final consumption, and is therefore influenced by consumer mobility; the geographic definition of a market is a central element in the evaluation of the competitive consequences of horizontal mergers; local transportation policies change the relative accessibility of different areas, and this may matter for local employment differentially across sectors; and the elasticity of local sales tax revenues to tax rates depends in part on the geographic substitutability of alternative sellers in the eyes of consumers.

In this paper, we provide the first large-scale, cross-sectoral description of geographical consumption markets, using more than 1.7 million credit card transactions by 70 thousand individual American consumers. We present evidence that consumers actively manage the spatial dimension of their purchases and introduce a simple framework motivating why consumers' optimal inventory and travel behavior may have implications for local equilibrium outcomes. We suggest durability/storability of a sector's output as a new characteristic that affects consumers' decisions, firms' responses, and observed gravity for a significant portion of economic activity. We propose the sector's average frequency of transactions as a proxy for this new characteristic. Using a separate data source, we then present evidence that consumer mobility influences local employment, establishment density, and average establishment size differentially across sectors.

We start by exploring overall consumer mobility and find it very limited: agents typically source purchases from just a few of the many locations available to them. Moreover, as broadly documented for merchandise trade both at international and intranational levels, gravity is a first-order feature of the data. Comparing consumers' behavior in their town of residence vs. outside ("at-home" vs. "out-of-home"), we find large drops in expenditures at very short distances: for the median sector, the expenditure in the average out-of-home location is only 34% of the expenditure in the home location. Comparing out-of-home expenditure at different distances, we find gravity present but weaker than other phenomena affected by spatial frictions: for the median sector, a 1% increase in distance decreases total expenditure by 0.4%, an elasticity substantially smaller than firm-to-firm intranational trade (-1.3%) or commuting flows (-4.4%).[1]

To understand the sources of the decline in expenditure with distance, we decompose total expenditure into contributions from 1) the total number of accounts (an "account" extensive margin), 2) the average number of transactions per account (a "frequency" extensive margin), and 3) the average value of a

[1] See for example Hillberry and Hummels (2007), and Monte, Redding and Rossi-Hansberg (2018).

transaction (a "batch size" intensive margin). We analyze the spatial decay of each component separately and by sector. In all cases, the extensive margin accounts for almost the totality of the declines. We find significant heterogeneity across industries in the overall impact of distance, that largely reflects the importance of the frequency margin. Moreover, we find stronger gravity in an industry correlated with a higher average frequency of transactions.

The correlation between gravity and frequency makes economic sense. If the cost of a person's trip is independent from the volume purchased, differences in the durability (or storability) of goods and services plausibly produce differences in spatial purchasing patterns across sectors: when goods are less storable/durable, consumers will want to buy smaller batches per trip, and buy more frequently; since travel is expensive, they will also buy closer to home. We think of storability/durability as a general characteristic of the sector, capturing the length of time over which a good or service can deliver its utility flow. For perishable items this concept is intuitive: fruit, for example, will depreciate if not eaten quickly. For durable goods, this concept may reflect depreciation due to use: consumers can store shirts for future use, and those shirts can deliver a utility flow for longer periods of time. For services, we think of storability/durability in a more general sense: while a consumer cannot buy two haircuts and store one for future use, she also does not visit the hairdresser multiple times a day; a haircut will "depreciate" over time more or less slowly as other goods do. While economists have long studied how exogenous characteristics of an industry (input-intensity or the weight-to-value ratio) have influenced trade patterns, most research has implicitly assumed a binary representation of durability/storability: goods are durable/storable and services are not; thus, goods are tradable and services are not. By examining expenditure patterns at a small geographical scale, we are able to uncover a more nuanced picture of tradeability. Hereafter, when we refer to "storability" and "storage costs" for brevity, we keep this more general interpretation in mind.

The fact that expenditure declines faster with distance in sectors that are transacted more frequently is suggestive of an active consumer role in determining local outcomes. However, other supply-side characteristics, like sectoral variation in producers' fixed costs, might influence store density in equilibrium, thereby affecting both how far consumers have to travel and how frequently they do so. To provide more direct evidence about consumer choice, we turn to two analyses at the consumer-account level. First, we show that higher-income consumers tend to have relatively fewer out-of-home transactions in sectors that are purchased more frequently at the national level. These findings hold even when we control for residence and individual, unobserved, time-invariant characteristics. If, as we argue below, the average frequency of transactions is a proxy for the durability/storability of a sector, this finding suggests that high-income individuals choose to travel shorter distances for the same sector. Second, we compare the behavior of the same individual under two potentially different travel-cost regimes using exogenous variation induced by rainfall. We assemble daily precipitation data from thousands of weather stations across the United States to examine patterns of the spatial distribution of transactions across sectors on rainy vs. non-rainy days. If consumers don't choose how far and how frequently to travel, then a change in travel costs common across sectors might impact the overall number of transactions, but would be less likely to affect their spatial distribution. For a given consumer, however, we find that rain affects the spatial distribution of transactions relatively less in sectors where the average frequency of transactions is higher. Since the supply network is fixed in the two travel-cost regimes, the difference is likely to come from an active choice by consumers, either via intertemporal substitution of trips, or via a choice of where to travel.

If consumers actively choose how far and how frequently to travel, and this choice is at least partially related to the storability/durability characteristics a sector's output, it is natural to expect that merchant employment and location choices take those consumer decisions into account. We then examine whether the spatial distribution of economic activity is affected by consumers' spatial choices in cross–sections of county data. First, we build a stylized model of shopping in space. Consumers are heterogeneous in travel costs, live on a point, consume a fixed quantity of a good per unit of time, and choose where to purchase it on a line. The good is consumed every instant at a constant rate, so consumers must hold an inventory to avoid infinitely frequent trips. Both travel and inventory holdings are expensive: hence, agents choose how frequently to buy the good, how much to buy per trip, and at which distance. For a given schedule describing price as a function of distance, consumer optimization delivers a spatial distribution of demand. Production can occur at any point on the line. Firms use land – a fixed factor – and labor. Perfect competition and free entry imply that at any location, the price is equal to the marginal cost of production, and all profits accrue to (absentee) landlords. A given price schedule implies a profit–maximizing spatial distribution of supply. The equilibrium price function makes supply and demand equal at every point, and determines a marginal plot of land where production no longer occurs.

Our simple model suggests that the frequency of transactions should be a proxy for storage costs, and generates a negative relation between gravity and frequency: when storability or durability is lower, consumers want to take more frequent trips, but want to travel closer to home; hence, expenditure in equilibrium declines faster in sectors where consumers take more frequent trips. If population increases, firms must proportionately increase output. Since travel is expensive, however, firms have an incentive to limit the expansion of the marginal plot of land used, and increase output per plot of land instead. When storage costs are large, this incentive intensifies: in response to the same increase in population, employment grows faster and the average distance between consumers and output grows more slowly in sectors with higher storage costs.

We explore these ideas empirically using the average observed frequency of transactions as a simple proxy for storage costs. We use the underlying geological composition of a county[2] to circumvent the endogeneity problem arising from regressing county–sector outcomes on population within the county. We find that in sectors where storage costs are higher, local employment grows faster in response to (exogenous) differences in population; moreover, the differential growth in employment is driven by the addition of establishments at a faster rate (i.e., increase in density), while employment per establishment grows at a slower rate. Our stylized model does not have explicit predictions for store density vs. employee-store size; however, our empirical findings are consistent with a more geographically concentrated demand arising from the need to save on travel time, and supply responding with a relative reduction in the

[2]See Burchfield, Overman, Puga and Turner (2006) and Duranton and Turner (2017), and our discussion in Section 6.

average distance between consumers and stores. Our results appear robust to a number of considerations, including among others the presence of labor supply feedbacks and shopping-near-work arising from commuting flows.

The rest of the paper is organized as follows. We relate our results to the existing literature in Section 2. Section 3 presents a set of cross-sectoral stylized facts. Section 4 presents evidence on individual behavior as a function of demographic characteristics and travel costs shocks. Section 5 introduces a highly stylized model to describe local equilibrium implications of differences in sectoral storage costs. Section 6 leverages those intuitions and presents evidence that consumer behavior matters for local equilibrium outcomes. Section 7 offers robustness analysis and comments on limitations of our analysis. Section 8 concludes.

2 Related Literature and Contributions

We relate our paper mainly to three areas of research and policy: spatial frictions in consumption, mergers' evaluation, and the role of transactions technology; we conclude the section with a discussion of other relevant literature.

Spatial frictions in consumption. While spatial analyses of the manufacturing sector abound, they carry limited information about final consumer behavior: intranational surveys typically record only firm-to-firm transactions, and only of merchandise. By focusing on final consumers, our work borrows from and complements the vast literature on spatial frictions (Anderson, 1979; Anderson and Van Wincoop, 2003; Eaton and Kortum, 2003; Hummels and Klenow, 2005; Hillberry and Hummels, 2008; Karádi and Koren, 2017, Gervais and Jensen, 2018).

The limited-but-growing literature on spatial consumption markets mostly focuses on specific industries, like restaurants (Couture, 2016, Davis, Dingel, Monras and Morales, 2018), food distribution (Handbury, Rahkovsky and Schnell, 2018), gasoline (Houde, 2012), health care (Raval and Rosenbaum, 2018) or movie theaters (Davis, 2006); studies focusing on consumption across cities, with less focus on consumer mobility, include for example Glaeser, Kolko and Saiz (2000), and Schiff (2015). Overall, we contribute to this literature by providing evidence on the nature and consequences of consumer mobility which are comparable across industries. We also extend the set of industries for which we can assess consumer mobility, and exploit cross-industry variation to argue that the storability of a sector's output is effectively a determinant of gravity and of local outcomes.

Our findings relate to a large portion of economic activity: final consumption accounted for around 70% of GDP in 2015 in the United States, and the service industries involved in its delivery accounted for around 40% of employment, and almost one-fourth of total value added. Understanding the nature of spatial patterns of consumption is essential for a wide spectrum of issues including the degree of spatial competition between firms; the determination of local and aggregate productivity and factors' income; the consequences of local labor demand shocks, local taxes and regulation; and the impact of investment in transportation infrastructure and of other "place-based" policies (Kline and Moretti, 2014). Our paper also provides useful information to evaluate the local employment consequences of the liberalization of international trade and investment in services: firms entering a foreign market decide the number and location of establishments based, among other things, on the strength of consumers' spatial frictions.

The geographic scope of a market. Our findings suggest a local empirical analogue of the classic proximity–concentration tradeoff (PCT) in foreign direct investment (FDI) (see for example Horstmann and Markusen, 1992; Brainard, 1997). Firms in a location choose whether to serve customers in a neighboring location by expecting them to travel ("export" in PCT) or by opening a new establishment closer to them ("FDI" in PCT); in sectors with high storage costs (high "transportation costs" in PCT), demand is more localized, and opening a new establishment, i.e. the "FDI" option, becomes more attractive.

In this sense, our results relate to the techniques and types of evidence on which the Department of Justice (DoJ) and the Federal Trade Commission (FTC) rely in assessing the competitive consequences of horizontal mergers (DoJ, 2010). The definition of a market plays a central role in the Horizontal Merger Guidelines of the U.S. DoJ and FTC: the market definition, among other things, "helps specify the section of the country where the competitive concern arises" and allows "to identify market participants and measure market shares and market concentration" (DoJ 2010, p. 13). The market definition always has a geographic component.

Our evidence is in line with these Agencies' indication that the geographic dimension of a market should account for spatial frictions that impede customers' travel. Further, we identify a particular sectoral characteristic, durability/storability, that interacts with consumers' mobility frictions to seemingly influence the spatial distribution of demand and the suppliers' endogenous geographical response. Our results may therefore provide future insights into key components of the evaluation of potential mergers indicated in DoJ (2010), like the identification of locations which are potential substitutes to two merging firms' sales, the related construction of market shares, and the assessment of a consumers' reaction to price increases.

Our findings also suggest that infrastructure and transportation policies, by modifying spatial frictions, may have heterogeneous consequences on current local competitive landscapes (Rossi-Hansberg, Sartre and Trachter, 2018); further, these policies arguably influence the desirability of future mergers in a way that is shaped by storability/durability.

Transaction technology. Our analysis relates to the role and consequences of transactions technology. The practical importance of direct sales from manufacturers to consumers, while fast growing, is still minor: for example, e-commerce sales (which include sales from a company's website and indirect sales from other distributors) account for only 6.4% of total retail sales in 2014, and only 0.9% in 2000, which is closer to our sample period, 2003 (Hortaçsu and Syverson, 2015). The literature about on-line transactions is growing and includes for example papers on the role of taxes in determining sales of on-line versus traditional retailers (Ellison and Ellison, 2009), the importance of distance and the persistence of home bias in on-line auctions (Hortaçsu, Martínez-Jerez and Douglas, 2009), or the gains from e-commerce (Einav et al., 2017). Our paper offers a detailed description of the nature and extent of spatial frictions that on-line commerce is trying to address; thus, it may provide intuitions on the exposure of different sectors to increased e-commerce presence.

Our work also implies that the purchasing technology available to optimizing consumers (the cost of a trip increasing in distance but fixed with respect to transaction size) may interact with heterogeneity in storage costs to generate measurable differences in local outcomes across sectors. In this sense, our analysis relates to Alessandria, Kaboski and Midrigan (2010), who show that the microstructure of firms' transaction technology has aggregate consequences on the level of trade, and to the broader literature on trade and storage costs (e.g. Williams and Wright, 1991, Coleman, 2009) and the demand for durable goods (e.g. Hendel and Nevo 2006a,b).

Other relevant literature. This paper is also related to several other strands of research.

First, our results are reminiscent of intuitions that can be traced back to Von Thunen's model of rural land use.[3] In his theory, a central city is surrounded by (concentric circles of) rural land with different possible uses. The transportation costs to the central market and perishability of different agricultural products determine how far from the city different products are produced: more perishable products like dairy (which may have higher storage costs) should be produced closer to the market.[4]

We complement and generalize some of the findings of the literature on cross-border consumption behavior (Chandra, Head, and Tappata, 2014, Agarwal, Marwell and McGranahan, 2017, Baker, Johnson and Kueng 2018); in particular, we study one source of cross-sectoral differences in the strength of gravity and its relationship to the frequency of transactions in the determination of consumer mobility and its employment consequences.

Part of the supply side in our analysis, the retail sector, is also subject of a growing literature: Jarmin, Klimek and Miranda (2005) and Hortaçsu and Syverson (2015) present some overall trends in the industry; Nakamura (2008) illustrates the role of the retail sector in price determination. Our work complements this literature by emphasizing the nature on consequences of spatial frictions in determining local demand.

Finally, Mian and Sufi (2014) study the response of broad tradeable and non-tradeable employment across U.S. counties to household balance sheets' shocks during the Great Recession; as a complement, we look at how employment in narrower sectors responds in the long run to sectoral differences in consumer mobility.

3 The Geography of Consumption

3.1 Data Description

We use a large proprietary dataset containing a sample of credit card transactions from a major financial institution. These transactions occurred roughly between March and October 2003. A transaction record contains, among other things, an exact date, an account ID, the amount spent, a Merchant Category Code (MCC – we will refer to it as a "sector") and information (to be processed) on the location of the merchant. In addition to all distinct transactions, we have information on the account itself, including the associated

[3] See for example Von Thunen (1966) for his 1826 model. Karádi and Koren (2017) explore a more modern version of the same argument. For an empirical test of the Von Thunen model, see for example Fafchamps and Shilpi (2003) about Nepal.

[4] Our work also relates to Central Place Theory (e.g. Christaller, 1933), that suggests a relation between city size and number of varieties available in a market. Our interest here is on consumer mobility and its impact across sectors, and we see our paper as complementary to this strand of literature. For an empirical test of Central Place Theory, see Schiff (2015).

ZIP code.[5] After cleaning the data, we have 1,722,873 transactions for 71,377 accounts (see Appendix A for a complete description of data cleaning and processing). The average transaction is 68 dollars, and total purchases amount to around $116 million. Table 1 gives a breakdown by 21 broad categories. The largest categories in terms of observations are Gasoline Services, Food Stores, Miscellaneous Retail, and Eating and Drinking Places.[6]

In the remainder of this section we show that distance affects consumer mobility heterogeneously across sectors, and that this heterogeneity is associated with how frequently the typical consumer makes a purchase. In the following section, we will present evidence that this association is driven at least in part by individual–level behavior.

Table 1: **Summary of transaction amounts (in USD), by sector**

Broad Category	**Median**	**Mean**	**St. Dev.**	**Sum**	**N**
Agricultural Services	83	136	212	1,307,616	9,615
Amusement, Rec. Serv.	45	89	169	1,771,086	19,897
Apparel	49	75	114	6,112,646	81,778
Auto Repair/Service/Parking	41	151	325	3,464,626	22,990
Auto and Truck Sales/Service/Parts	66	198	423	6,624,392	33,473
Building Mat./Hardware/Garden Supp.	42	101	258	9,658,412	95,568
Communications	53	91	122	559,201	6,113
Durable Goods	68	209	521	837,246	4,004
Eating and Drinking Places	26	39	73	8,770,958	227,715
Food Stores	30	46	59	12,116,604	265,828
Furniture, Home Furnishings, Equip.	60	194	430	10,853,963	55,917
Gasoline Services	19	22	31	6,934,785	312,670
General Merchandise Stores	43	67	122	13,963,544	207,866
Health Services	71	164	375	4,487,799	27,381
Hospitality	96	170	308	6,430,175	37,934
Misc. Retail	32	65	182	16,100,792	248,069
Misc. Services	95	316	703	1,870,560	5,919
Motion Pictures	14	19	44	272,948	14,048
NonDurable Goods	38	78	175	640,203	8,246
Other Vehicles Sales/Service/Parts	76	259	746	1,366,449	5,279
Personal Services	37	74	210	2,406,679	32,563
Total	30	68	188	116,550,684	1,722,873

3.2 Consumers Visit Few Locations

We start our exploration by considering how far consumers travel across locations for purchases. A "location" in the data is identified at the level of Census incorporated place or county subdivision. In the

[5] The same original source data was used in Agarwal, Marwell, and McGranahan (2017).

[6] Table C.1 in the Appendix (page 58) shows summary statistics by state of purchase. The largest number of transactions are reported in New York, California, and Massachusetts.

raw data with all transactions – including those of tourists living in one place and spending in another, potentially far away place – we count expenditure flows between 17,535 unique locations (11,454 unique residence and 14,962 unique sale locations). The data records transactions among 232,927 unique pairs. There are 7.4 transactions per pair, and the median pair has 1 transaction. Naturally, it is unrealistic to assume that very long distances reflect day-to-day consumption behavior: for example, only about 4.3 million potential location pairs (1.4% of all potential pairs) in our data have distance below 120 km.[7] Among those pairs, around 1.5 million transactions are recorded between only 120,783 pairs, or 2.8% of the possible pairs, with 12.3 transactions per pair on average, and 2 transactions for the median pair. Overall, the matrix of residence-sales location purchases is sparse.

To dig deeper into this low mobility, we focus in Table 2 on all transactions occurring within 120 km and construct a residence location–level dataset. For each residence, we compute the total number of locations visited by all consumers living in that residence, and the total number of locations in the data within 120 km. We also compute the average distance from the residence location to these locations. Table 2 shows summary statistics on the distribution of locations visited. The first row shows that consumers in the median residence visit only 7 distinct sales locations during the sample period (11.2 sales locations on average). One might think that this low number is simply a consequence of the absence of close-by options, but this is not the case: the second row in the table shows that consumers living in the median residence have 192 sales locations within 120 km. The third row shows that there is indeed a large variation in the average distance of different residence locations relative to their shopping options. The overall fraction of available locations where purchases actually occur (fourth row) is very small.

Table 2: **Summary statistics across residence locations (transactions within 120km)**

variable	**min**	**p10**	**p25**	**p50**	**p75**	**p90**	**max**	**mean**	**N**
Sales locations visited	1	2	3	7	14	27	445	11.17	9,479
Sales locations available	2	66	112	192	338	646	1,115	271.82	9,479
Mean distance to sales locations	16.8	63.1	69.7	76.1	80.6	83.6	97.2	74.41	9,479
Share available locations visited	0	0.01	0.02	0.04	0.07	0.12	0.67	0.05	9,479

In Table 3 , we ask what accounts for this low mobility. In column 1 of Table 3 we regress the log of number of sales locations visited on the log of number of sales locations available and find an elasticity of 0.55: overall, the number of visited locations grows at about half the pace of the available locations. Distance, on the other hand, has a stronger role: controlling for the number of available sales locations, a 1% increase in average distance to those locations is accompanied by a 2.4-2.6% decrease in the number of locations visited (columns 2 and 3).[8] These results suggest a central role of distance on consumer

[7] Distance is always computed between the centroids of two locations using the Haversine formula. When looking at the impact of distance on flows below, we will also restrict our attention to transactions with distance up to 120km. Monte, Redding, and Rossi-Hansberg (2018) find this threshold to be one where gravity in home-to-work commuting flows has a change in slope, so it is a natural cutoff.

[8] In Tables C.2 and C.3 of Appendix C.2 (page 57), we repeat this analysis using a sample of users with at least one transaction every two days. The fraction of locations visited has a similar distribution. In a regression analysis, distance has twice the impact of number of available locations on the total number of locations visited.

expenditure. We explore this aspect next.

Table 3: Locations available and locations visited

Dependent variable:	Log of number of sales locations visited		
	(1)	(2)	(3)
Sales locations within 120km, log	0.548***		0.568***
	(0.010)		(0.010)
Average distance to sales locations within 120km, log		-2.374***	-2.594***
		(0.125)	(0.090)
Constant	-0.988***	12.096***	10.061***
	(0.052)	(0.539)	(0.399)
R^2	0.22	0.09	0.32
N	9,479	9,479	9,479

Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

3.3 The Distance Traveled Varies by Sector

This first snapshot paints a picture where consumers have many options but choose to shop only in a limited number of locations, and a strong role is played by distance. How far, then, do people travel for their purchases?

The median transaction in the data occurs about 9 km from home. There is large dispersion around this typical value: the first 25% of transactions occur within the same place, while the third quartile is around 30 km. A long right tail of high distances is likely due to account holders traveling outside town for work or tourism.[9] While these and other details are relegated to the Appendix, we show in Figure 1 select percentiles of the distances at which transactions occur, by sector.[10] The heterogeneity in distance traveled is very significant: moving from a sector at the 10th percentile to a sector at the 90th, the median distance traveled goes up by a factor of around 7. The patterns make sense overall: the median transaction occurs at 4 km for staple items like Food Stores, and around 12 km for Eating and Drinking Places; it is, however, above 20 km for Durable Goods and 33 km for Amusement and Recreational Services, which are likely purchased less frequently.[11]

[9] Online transactions have been eliminated as much as possible. See the Data Processing section in the Appendix for more details.

[10] Tables C.4 and C.5, in pages 60 and 61 respectively, show percentiles in the distribution of transaction distances by sector in the raw data and weighted by value of the transaction. The typical dollar is spent farther than where the typical transaction occurs, as reflected in right-ward shifts in the value-weighted distributions.

[11] Interestingly, Davis (2006) finds that larger population within 10 miles increases demand to a movie theater, and that the geographical market of a theater extends for at most 15 miles around it: we find for the same industry that 75% of the transactions occur in fact within (around) 11 miles.

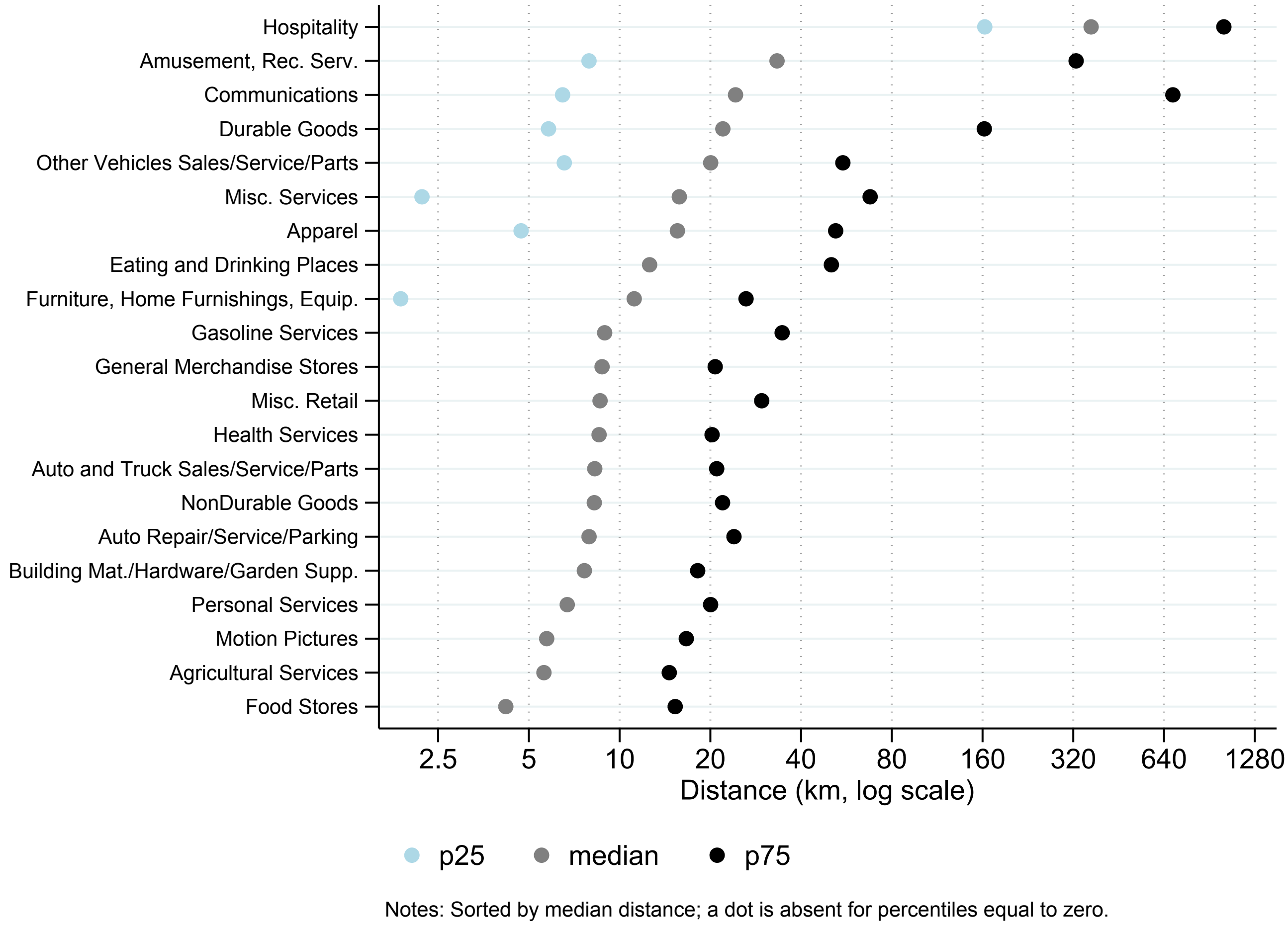

Figure 1: Distances traveled by sector (select percentiles)

Obviously, the distance traveled by consumers is a combination of their willingness to travel (as mediated by their optimal shopping behavior) and supply conditions like the density of producers. We will return to this distinction later. For now, we emphasize that the spatial dimension of consumer behavior is actively moving in the data: consumers visit just a few locations among the many available, but the typical distance traveled varies broadly across sectors. To understand more the local nature of different consumption markets, we need to explore further the determinants of the relation between total purchases and distance. We move to this task next.

3.4 Gravity in Consumer Expenditure

Gravity is an almost universal feature of spatial relationships. While a substantial amount of literature has documented the decay of goods' trade flows with distance at international and intranational levels, little is known about the spatial behavior of final consumers.[12] We fill this gap in two steps.

First, we document that gravity also holds for consumers' behavior.[13] We make full use of the information available in the data comparing 1) expenditure inside vs. outside one's place of residence, and 2) the decline in expenditure across merchants at different distances from the residence location.

We then analyze the margins of this decline, decomposing the total decay into the number of accounts transacting (an extensive "accounts" margin), the number of transactions per account (a "frequency" margin), and the average expenditure per transaction (a "batch size" intensive margin).

3.4.1 Expenditure patterns display gravity

We start our exploration of gravity by investigating how quickly total expenditure decays with distance. A large empirical literature has documented that merchandise trade flows decay with distance both across countries (e.g., Disdier and Head 2008) and within countries (e.g., Hillberry and Hummels 2007). Since final consumers buy goods directly from producers in relatively few cases, our knowledge of gravity in final consumption is extremely limited. Moreover, virtually all the literature deals with merchandise shipments, ignoring the service sector altogether. Here, we fill these important gaps.

For a given sector, we denote with X_{hs} the total expenditure of accounts residing in location h on merchants selling in location s. Except where noted otherwise, from now on we restrict the data and use only transactions occurring at distances up to 120 km. We initially relate the expenditure X_{hs} to the distance between residence and merchant locations in two ways. First, we simply estimate the change in expenditure associated with shopping out of the home residence ("out-of-home"):

$$\log X_{hs} = \alpha + \gamma^{(h)} + \gamma^{(s)} + \eta \times \mathbf{1}_{(h \neq s)} + \varepsilon_{hs} \tag{1}$$

where $\mathbf{1}_{(h \neq s)}$ is an indicator function assuming the value of 1 if $h \neq s$ and zero otherwise; in this regression, the expenditure of residents on their home location, X_{hh}, is included. Second, we follow the gravity literature and estimate the impact of distance on trade flows with a regression of the form

$$\log X_{hs} = \alpha + \gamma^{(h)} + \gamma^{(s)} + \delta \log dist_{hs} + \varepsilon_{hs} \tag{2}$$

In this equation, α is a constant, and $dist_{hs}$ is the distance between the centroids of h and s; this regression includes only pairs where $h \neq s$, since $dist_{hh} = 0$. In both equations, α is a constant, and a set of origin and destination fixed effects, $\gamma^{(h)}$ and $\gamma^{(s)}$, controls for unobserved differences in size, productivity and intensity of competition (Anderson and Van Wincoop, 2003). These two approaches highlight complementary features of the data. The coefficient η in Equation (1) measures the expenditure drop associated with visiting the average out-of-home location; hence, it shows the importance of very

[12] International flows of goods are only measured at the country level, thus ignoring the travel dimension of consumers' purchases. Intranational flows of goods typically record firm-to-firm transactions.

[13] To our knowledge, the first formulation of a gravity law for final consumers was proposed by Reilly (1931): "Two cities attract retail trade from any intermediate city or town in the vicinity of the breaking point, approximately in direct proportion to the populations of the two cities and in inverse proportion to the square of the distance from these two cities to the intermediate town."

short trips, for which distance is poorly measured. The coefficient δ in Equation (2) shows the elasticity of expenditure to distance considering only out–of–home expenditure flows, in which case distance can be measured.

We first estimate Equations (1) and (2) across all sectors. We find, unsurprisingly, very clear distance effects. Estimating (1), the expenditure in the average location out-of-home is only about 8.8% of the average expenditure at home ($\eta = -2.435$, robust s.e. 0.021).[14] When we estimate (2), we find a slope of -1.051 (robust s.e. of 0.006), in line with estimates in the trade literature.[15] A comparison of these two coefficients shows that a large decay already appears at very short distances.

These pooled estimates mask large differences across sectors. Table 4 shows the coefficients of η (column 1) and δ (column 4) when we estimate Equations (1) and (2) by sector.[16] Sectors in this table are ordered by the out-of-home dummy in column 1 (this ordering will be kept throughout the paper for ease of reference). The strong decay at short distances is pervasive across sectors. However, the decay is heterogeneous: in sectors like Food Stores, the expenditure in the average location out-of-home is around 9% of the expenditure at home; this fraction grows to 20% for Eating and Drinking Places, 38% for Personal Services, and at 91% for Durable Goods.

The impact of distance as measured by estimates of (2) is consistent with this picture: the correlation between the two sets of coefficients across sectors is 0.68. However the much smaller distance coefficients in these estimates are notable: for the typical sector, a 1% increase in distance is associated with a 0.41% decrease in expenditure, and almost all coefficients are below the benchmark value of approximately 1 for international trade. We conclude that most of the decline in expenditure happens at short distances as measured in (1), which we will focus on for much of the remaining analysis.

The final column of Table 4 reports, for each sector, the simple average of the number of purchases per account in the transaction data. We note for now the tendency of sectors with stronger gravity to be purchased more frequently. We will return to this in more detail after we have a better understanding of the margins of decay accounting for the expenditure drop over space.

3.4.2 Margins of adjustment

Why does expenditure decay with space? As distance increases, there may be fewer people traveling out-of-residence; moreover, those who are traveling may do so less frequently, or spend a different amount per transaction. These margins map into simple decompositions in the spirit of Hummels and Klenow (2005) and Hillberry and Hummels (2007). In any given sector, we express total expenditure of consumers in h falling on merchants in s as

[14] Using all data, we find $\eta = -2.545$ (robust s.e. 0.0223).

[15] This slope is not particularly sensitive to changes in the cutoff. See Appendix C.4, page 62, for further discussion.

[16] All p-values are computed using heteroskedasticity–robust standard errors. The number of observations reported excludes "singletons", i.e. those observations that would be absorbed by fixed effects and do not contribute to the estimation.

Table 4: **Decline in expenditure**

Category	Out of Home coeff	Out of Home pv	Out of Home obs.	Gravity coeff	Gravity pv	Gravity obs.	Frequency of transactions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Food Stores	-2.23	0.00	22,649	-0.85	0.00	18,632	7.53
Gasoline Services	-2.08	0.00	39,666	-0.60	0.00	34,615	8.86
General Merchandise Stores	-1.78	0.00	26,837	-0.93	0.00	23,932	5.14
Misc. Retail	-1.70	0.00	34,052	-0.65	0.00	30,042	5.25
Eating and Drinking Places	-1.57	0.00	34,504	-0.56	0.00	31,022	5.93
Building Mat./Hardware/Garden Supp.	-1.40	0.00	14,185	-0.73	0.00	11,604	4.15
Auto Repair/Service/Parking	-1.25	0.00	4,414	-0.40	0.00	3,013	1.83
NonDurable Goods	-1.16	0.00	978	-0.65	0.00	758	1.68
Health Services	-1.12	0.00	5,134	-0.33	0.00	3,910	2.16
Apparel	-1.10	0.00	15,918	-0.53	0.00	14,066	2.91
Furniture, Home Furnishings, Equip.	-1.07	0.00	12,286	-0.57	0.00	10,734	2.33
Auto and Truck Sales/Service/Parts	-1.04	0.00	7,298	-0.33	0.00	5,508	1.98
Motion Pictures	-1.04	0.00	1,922	-0.34	0.00	1,248	2.16
Amusement, Rec. Serv.	-1.03	0.00	2,958	-0.23	0.00	2,329	2.03
Personal Services	-0.96	0.00	5,203	-0.31	0.00	3,760	2.46
Misc. Services	-0.92	0.06	220	0.91	0.02	116	1.57
Communications	-0.89	0.00	424	-0.41	0.01	263	1.36
Agricultural Services	-0.88	0.00	552	0.42	0.11	190	1.86
Other Vehicles Sales/Service/Parts	-0.68	0.41	257	-0.59	0.08	128	1.64
Hospitality	-0.64	0.01	1,392	-0.14	0.08	1,158	1.53
Durable Goods	-0.09	0.90	79	1.11	0.67	15	1.64

$$X_{hs} = \underbrace{N_{hs}}_{\text{account margin}} \times \underbrace{\bar{x}_{hs}}_{\text{expenditure margin}} = \tag{3}$$

$$= \underbrace{N_{hs}}_{\text{account margin}} \times \underbrace{f_{hs}}_{\text{frequency margin}} \times \underbrace{\bar{x}_{hs}/f_{hs}}_{\text{batch size margin}} \tag{4}$$

Equation (3) says that as distance increases, expenditure can decrease either because the number of agents traveling decreases (the extensive "account" margin) or because agents spend less on average. In turn (4) suggests that lower expenditure per account on average can arise either because each transaction is smaller (the "batch size" margin) or because consumers transact less often (the "frequency" margin). When we re-estimate Equation (1) with the left side being each of these three terms, the coefficients on the out-of-home dummy add up to the overall coefficient η reported in column 1 of Table 4 (and similarly for Equation (2)).[17]

[17] A further angle of this decomposition could relate to the Alchian and Allen (1964) conjecture: consumers should be willing to travel more for higher quality goods and services when travel costs do not vary with quality. Hence, there should be a positive relation between average value of a transaction and distance. Unfortunately, our data does not allow a precise measurement of unit values and hence cannot be used to speak to this conjecture. For related work on international trade, see Hummels and Skiba (2004).

Figure 2 shows the results of this decomposition for (1). The length of each bar corresponds to column 1 in Table 4. We find two broad messages.

First, most of the decline in expenditure over space is due to fewer people traveling outside, or people taking less frequent trips. The blue bar measures the contribution of the "account" margin. For the typical sector, 72% of the drop in out-of-home expenditure is associated with fewer people traveling outside, rather than to people spending less on average for out-of-home transactions.[18] As a benchmark, Hillberry and Hummels (2007) find, for firm-to-firm shipments within U.S., that for short distances the extensive margin explains almost the totality of the decay.

The remaining part of each bar measures the decline due to lower average expenditure per account. The gray section indicates that the average expenditure per account drops outside of home almost exclusively because of the "frequency" margin: consumers spend less on average out-of-home because they choose to travel outside less frequently, not because they spend less per transaction. The drop in the average transaction value (the "batch size" margin) has a limited role in most cases. Tables C.8 and C.9 in the Appendix (p. 64) show that the combination of the "account" and "frequency" margins typically contributes 90%-95% of the decline in expenditure.

Second, Figure 2 suggests that a large part of the heterogeneity in gravity seems associated with heterogeneity in the frequency margin and not associated with the "batch size" margin – i.e., the length of the bar varies because of variation in the gray section.[19] This is very apparent when we plot the out-of-home expenditure as a share of home expenditure $\exp(\eta)$ (using column 1 in Table 4) against the average number of transactions per account in the sector from the data. Figure 3 shows this relation for the sectors where the out-of-home dummy is statistically significant from zero at 10% level. A simple regression line through this data has a slope of -.69 (robust s.e. 0.07) and an R^2 of 0.86.[20] When customers choose more visits, gravity is more important. Note that since the average number of transactions has not been used directly to compute the out-of-home dummy, there is nothing mechanical about this empirical relation.

Our stylized model below will provide a possible explanation for this correlation, based on heterogeneity in the storability/durability of the sector. When storage costs are high, consumers want to reduce the average inventory held. To do so, they need to purchase smaller batches, but more frequently. Since travel is expensive, however, a higher frequency can only be optimal with reduced distances. Hence, across sectors, if storage costs are higher, the frequency of purchase should grow but the expenditure should decline faster with space. This behavior generates a negative correlation between the strength of gravity and frequency of transactions, as present in Figure 3.

These results provide the first piece of (somewhat indirect) evidence that demand conditions may

[18] Tables C.6 and C.7, in the Appendix (p. 63), show the actual values of the "account" and "expenditure" margins with associated p-values for both Equations (1) and (2).

[19] A simple regression of the out-of-home dummy on each of the "account", "frequency", and "batch size" margin coefficients separately has R^2 of 75%, 87%, and 7% respectively.

[20] The figure excludes Durable Goods and Other Vehicles Sales/Service/Parts, an outlier. Using all estimates, the regression line would have a slope of -0.77 (robust s.e. 0.09), with $R^2 = 0.78$. Figures C.2 and C.3 in the Appendix, starting at page 66, replicate this figure for all the sectors and for the impact of distance using eq. (2). We have also experimented with an alternative measure of frequency that gives more weight to users which spends more overall in the data, with very similar results.

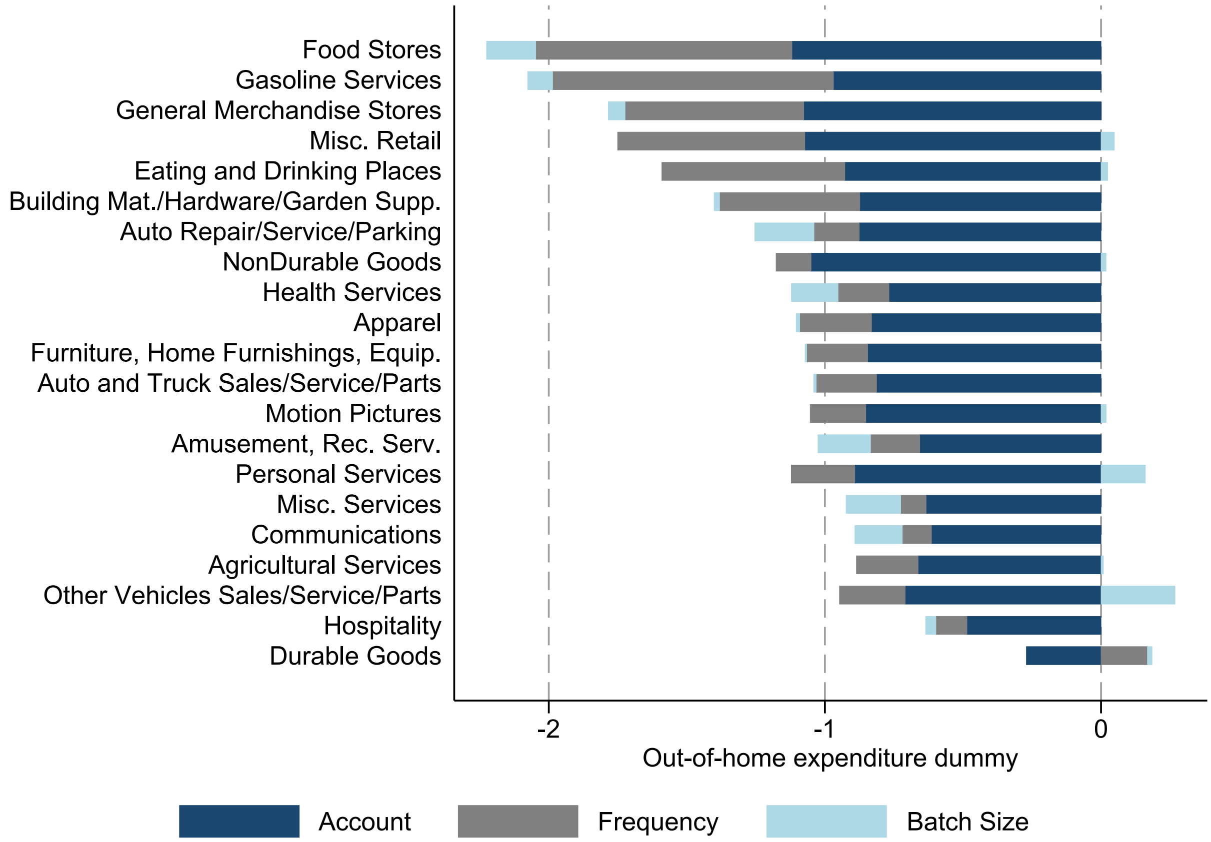

Figure 2: Margins in the out-of-home expenditure drop

matter in determining local equilibrium outcomes. In sectors where storage costs are high, consumer demand declines faster with distance and demand is more spatially concentrated. Hence, suppliers should be willing to increase production in larger markets by conserving consumer travel time.

In the remainder of the paper, we develop these arguments in more detail. We first show that the cross–sectional relation in Figure 2 is not purely driven by unobserved heterogeneity in supply characteristics. We then develop a simple partial equilibrium model to clarify how consumers' purchasing behavior can impact local equilibrium outcomes. Finally, we present reduced–form evidence that the average frequency of transactions, a proxy for storage costs, is related to heterogeneity in employment and store density across sectors in a way that is consistent with the mechanism we have described.

4 Individual Level Responses

The analysis up to now has shown that consumers' typical travel ranges are limited, expenditure declines with distance, and the combination of "account" and "frequency" margins are the main reasons for the decline. We have also shown that the strength of gravity varies by sector with the frequency of travel,

Figure 3: Drop in expenditure out of home

which we interpret as related to sector-level characteristics like storability and durability.

While certainly suggestive, these facts by themselves do not show that the consumer is actively managing the spatial dimension of consumption, that is, how far to travel; some sector-level characteristics on the supply side may bring producers and consumers closer to each other, so that the observed distance traveled is shorter. One such characteristic could be the fixed costs of operating a store, which can in principle affect the density of stores in equilibrium. Suppose, for example, that food stores have a low fixed cost of operation, so they are more densely distributed over the territory than, say, apparel stores. Suppose now that consumers do not choose how far to shop, but simply buy whenever they are close enough to a store. If consumers are more likely to take short trips than long trips, then we would observe more frequent transactions in food stores than in apparel stores, everything else equal, since food store are denser; as a consequence, gravity will also be stronger for food stores because expenditure will be closer to the consumer's residence than for apparel stores. This mechanism would replicate the correlation in Figure 3 without consumers choosing how far to travel for their purchases.

To make progress on this issue, we exploit individual-level variation in our data. Since our data comprises a relatively short time span, we can assume that the supply network is fixed. In a first exercise, we show that after accounting for individual-level heterogeneity, the propensity of individuals to travel out-of-residence for different sectors varies systematically with individual-level characteristics like income: as we move from low- to high-purchase frequency sectors, richer individuals tend to have relatively fewer transactions out-of-residence. In a second exercise, we identify for each consumer two different travel costs regimes by recovering whether it was raining or not on the day of a particular transaction. We find that while rain makes a given consumer less likely to travel out–of–residence for his purchases on average, this effect is relatively more muted for frequently purchased sectors. These two results suggest that for fixed supply network, consumers are actively choosing how far to travel.

4.1 The Role of Individual Characteristics

We are interested in whether the travel behavior of consumers in the same shopping environment varies across sectors with demographic characteristics. Since the number of transactions comes in integer values, a Poisson model is an appropriate starting point. In particular, we will estimate Poisson models where the mean number of out–of–residence transactions for individual i in sector s, $out_{i,s}$, takes the form:

$$E\left[out_{i,s}|x_i, \bar{f}_s, \delta^i, \delta^s\right] = \exp\left\{\alpha + \beta_0 \cdot n_{i,s} + \beta_1' \delta_i + \gamma_0 \cdot \delta^s + \gamma_1' \left(\bar{f}_s \cdot x_i\right)\right\} \tag{5}$$

In this expression, $n_{i,s}$ is the total number of transactions for account i in sector s; δ^i and δ^s are either individual i (sector s) characteristics or individual (sector) dummies; x_i are individual characteristics; $\bar{f}_s$ is the average frequency of transactions for sector s in the overall data. Since we are controlling for the total number of transactions, we are effectively examining the response of the geographical composition of purchase to different covariates. Moreover, the spatial aggregation of the geographical units of aggregation play a more limited role since our dependent variable sums across all locations other than the residence location. Econometrically, our coefficients of interest are the interactions of demographic characteristics with the average frequency of transactions.

Table 5 shows the results of this estimation.[21] In all specifications, we cluster standard errors at the individual level to allow for correlation among observations pertaining to the same individual.[22] All the models include sector-fixed effects. Column (1) shows the baseline elements of our regression. The expected number of transactions out-of-residence in the sample period increases by 1.6 percentage points for an individual with one additional transaction;[23] a 10% higher income increases the number

[21] To ensure our results are a faithful representation of individual behavior, we limit the analysis to "frequent users," i.e., users with at least 120 transactions in our sample. We further require these accounts to have valid (self-reported) income and age. These individual-level analyses are based on about 1,400 individual accounts. Since we have 21 sectors, our data comprises roughly 29 thousand observations.

[22] In other words, standard errors are immune to the equidispersion assumption in a strict Poisson maximum–likelihood estimation. In Appendix C.7 we report estimation results for a negative binomial model where overdispersion is explicitly taken into account, and discuss potential drawbacks of such an alternative. We have also experimented with alternative clusterings of the stadard errors: for example, we have split individuals in 5 income quintiles and 5 age quintiles; we then have formed indicators for the 525 combinations of age group times income group time sector and clustered our standard errors using this alternative categorization. The estimates of the interaction coefficients maintain the same patterns of significance.

[23] In these models, the total number of transactions appears in levels to keep all observations with zero transactions in the

of transactions out-of-residence by 1.39% on average.[24] The consequences of a higher income are not surprising: while higher income individuals have a higher opportunity cost of time, they also likely have access to better means of transportation. Age of the individual per se does not seem to affect the frequency of transactions out-of-residence.

Table 5: **The role of individual heterogeneity**

	(1)	(2)	(3)	(4)
Dependent Variable:	Number of transactions out of residence			
Number of transactions	0.016***	0.016***	0.022***	
	(0.001)	(0.001)	(0.001)	
Log age	-0.026	0.053	0.057	
	(0.059)	(0.108)	(0.125)	
Log income	0.139***	0.477***	0.255***	
	(0.029)	(0.048)	(0.057)	
Log age × log frequency of transactions		-0.045	-0.061	-0.043
		(0.060)	(0.056)	(0.074)
Log income × log frequency of transactions		-0.202***	-0.196***	-0.276***
		(0.029)	(0.025)	(0.031)
Observations	28,959	28,959	28,959	28,959
Sector fixed effects	Yes	Yes	Yes	Yes
ZIP code fixed effects	No	No	Yes	No
Individual fixed effects	No	No	No	Yes
Pseudo R-Square	.61	.61	.74	.64

Standard errors clustered at account level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

In the second column, we interact log age and log income with the average frequency of transaction in a sector. We find that higher-income individuals are relatively less likely to travel out-of-residence for more frequently purchased goods: fixing the total number of transactions, a 10% richer individual transacts 2% less out-of-residence for every one-point increase in the log-frequency of transactions. For example, the same movement from a low-frequency sector like Durable Goods (20th percentile of frequency, about 1.6 transactions in the sample on average), to a high-frequency sector like Miscellaneous Retail (80th percentile of frequency, about 5.3 transactions) reduces the number of transactions out-of-residence by $2.02 \cdot \ln(5.3/1.6) = 2.42\%$ more for a 10% richer individual, compared to a poorer one. Age is still statistically insignificantly associated to our dependent variable and stays so throughout this table.[25]

estimation sample.

[24] In models where the mean function is exponential and the dependent variable in levels, the coefficient on a regressor in levels is interpreted as a semi–elasticity and the coefficient on a regressor in logs as an elasticity.

[25] In all tables in Appendix C.7, we also control for interactions between economy–wide log average number of employees per store (a proxy for sector–level fixed costs) and log age and income to limit as much as possible the role of supply–side

The remaining two columns control for progressively more detailed sources of unobserved heterogeneity. In column (3), we introduce fixed-effects for the zip code residence location of the account-holder. This allows us to fix the shopping environment and in practice compare two individuals living in the same narrow area. The last column uses individual-fixed effects: here, we control for all time-invariant individual characteristics (e.g., wealth, education, precise residence location, overall use of credit cards).[26] It is always the case that, controlling for the overall number of transactions, higher-income individuals transact relatively more locally for frequently purchased sectors.

These results show that differences in purchasing behavior are significantly associated with individual characteristics like income: high-income individuals tend to shop relatively more locally in sectors which are purchased more frequently. This is consistent with a situation where prices close to home are higher than farther away, particularly in sectors where storage costs are high, and high-income people are relatively more willing to incur these extra costs. In general, these results support the view that agents actively choose how far to travel for their purchases.

The findings in this section are based on the average behavior of individuals over time. In the next section, we examine the response of the average individual to differences in travel cost regimes arising from weather conditions.

4.2 The Effect of Rain

In this section, we exploit a different source of plausible variation in individual behavior: weather conditions on the day of the transaction. If individuals are not actively choosing how far to travel for their purchases, then bad weather might impact the overall number of transactions but not individuals' propensity to purchase out-of-residence. Again, we are only interested in the spatial distribution of transactions.

To pursue this line of analysis, we turn to daily data on rainfall precipitation from the National Oceanic and Atmospheric Administration, as described in Menne et al. (2012). For each centroid of a residence location in our data, we find the closest weather station among the roughly twelve thousand disseminated across the United States. In the transaction data, the median distance between a weather station and a residence is 7.3 km (mean 8 km). We use this daily data on rainfall to assign a weather status for each transaction: we create a transaction-level indicator variable that assumes the value of 1 if, during the transaction day, the associated weather station recorded rain in the residence location. During the sample period, 34% of transactions have a rain episode so defined. A concern could be that most of the variation in this indicator is geographically related, rather than occurring within residence locations over time. This is not the case. A regression of the weather status indicator variable on residence-location fixed effects and transaction-date fixed effects absorbs only 17% of the variation in the transaction level data, leaving ample residual variation to identify movements in the propensity of purchase outside of one's residence.

density. The pattern of significance for our interaction variables remains unchanged. The sector–level correlation between log average frequency and log average number of employees per store is 0.12 (pvalue 0.6).

[26] While in general individual–level fixed effects may give rise to an incidental parameters problem, this is not the case for Poisson regressions (see for example Cameron and Trivedi, 2005).

We then construct an extended dataset, starting from the analysis in the subsection above, where for each individual we count $out_{i,s,r}$, the number of transactions out-of-residence by sector during rainy ($r = 1$) and non-rainy days ($r = 0$). Our interest is to understand if, conditioning on the total number of transactions in a sector, visits to out-of-residence stores are systematically related to differences in travel conditions, and whether this relation is heterogeneous across sectors. In particular, in this section we estimate Poisson models where the mean number of out–of–residence transactions for individual i in sector s under rain conditions r, $out_{i,s,r}$, takes the form:

$$E\left[out_{i,s,r}|x_i, \bar{f}_s, \delta^i, \delta^s, \delta^r\right] = \exp\left\{\alpha + \beta_0 \cdot n_{i,s,r} + \beta_1' \delta_i + \gamma_0 \cdot \delta^s + \gamma_1' \left(\bar{f}_s \cdot x_i\right) + \eta_0 \cdot \delta^r + \eta_1 \cdot \delta^r \bar{f}_s\right\}$$

In this expression, $n_{i,s,r}$ is the total number of transactions recorded for account i by sector s and weather status r; δ^r is a dummy equal to 1 for observations pertaining to rainy days and zero otherwise; and all the remaining notation follows Equation (5) above. The results of this analysis are reported in Table 6, which broadly mimics the structure of Table 5 above. Again, all standard errors are clustered at individual level.[27]

Table 6: **The effect of rain**

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Number of transactions out of residence				
Number of transactions	0.023***	0.023***	0.023***	0.035***	0.036***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Log age	-0.020	-0.019	0.079		
	(0.066)	(0.066)	(0.113)		
Log income	0.115***	0.115***	0.513***		
	(0.032)	(0.032)	(0.052)		
Rain dummy	-0.314***	-0.816***	-0.815***	-0.909***	-0.910***
	(0.019)	(0.038)	(0.038)	(0.040)	(0.041)
Rain dummy × log frequency of transactions		0.302***	0.302***	0.419***	0.421***
		(0.026)	(0.026)	(0.031)	(0.033)
Log age × log frequency of transactions			-0.055	-0.063	-0.062
			(0.067)	(0.057)	(0.057)
Log income × log frequency of transactions			-0.237***	-0.206***	-0.202***
			(0.036)	(0.025)	(0.025)
Observations	57,918	57,918	57,918	57,918	57,918
Sector fixed effects	Yes	Yes	Yes	Yes	Yes
ZIP code fixed effects	No	No	No	Yes	No
Individual fixed effects	No	No	No	No	Yes
Pseudo R-Square	.56	.56	.56	.68	.69

Standard errors clustered at account level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

[27] Again, we have experimented with clustering our standard errors at the rain dummy - sector level, without changes in the pattern of significance.

Column (1) reports the baseline model estimated without interactions, but including the dummy for transactions on rainy days. The dummy is negative and statistically significant: the average number of transactions out-of-residence on rainy days is $\exp\{-0.314\} = 0.73$ times the number of transactions out-of-residence on non–rainy days, for a given total number of transactions of the account-holder (and given demographic characteristics): intuitively, transactions are relatively more local when it rains. Column (2) shows, however, that the degree to which transactions become more local varies by sector. The interaction between the rain dummy and the frequency of transactions is large, positive and statistically significant: for example, the out-of-residence transactions on rainy days decline to 51% of the value on non–rainy days for durable goods, but only to 73% for miscellaneous retail; for gasoline, the most frequently purchased, the same number is 85%. The prevalence of local trips changes differentially across sectors. Column (3) again introduces interactions of log age and income. The coefficient on the rain $\times$ frequency does not change (up to rounding), and the coefficients for the interactions of income and age are very similar to those in Table 5 above.[28] Columns (4) and (5) add, progressively, residence– and account level fixed effects. As we do so, the heterogeneous impact of rain becomes stronger. In the most restrictive specification of column (5), for a given number of transactions, the frequency of out–of–residence transactions on rainy days declines to 50% of the value on non–rainy days for durable goods, but it is 81% of the non–rainy days value for miscellaneous retail and it is unchanged for gasoline.[29]

Taken together, the results of this section are consistent with the spatial distribution of transactions being an active margin of consumer choice. Conditioning on the shopping environment that an agent has available, this distribution varies with individual characteristics. Moreover, when the same consumer is exposed to different travel cost regimes the spatial distribution of transactions varies differentially across sectors with the frequency of transactions, a measure that we associate to characteristics of storability or durability of the sector itself.

While this exercise points to an active role of consumers, it is not designed to identify any long–run causal effect: hence, we cannot use it to understand whether optimal consumer behavior has an overall impact on local equilibrium outcomes. We tackle this broader question in Section 6. Before doing that, we develop a more formal association between frequency of transactions and storability, formalize why storability can drive a negative association between gravity and the frequency of transactions, and develop a simple intuition for the relation between storage costs and equilibrium outcomes.

5 A Simple Model of Shopping

We observed in Section 4 that consumers actively manage the spatial distribution of their transactions. However, an important question remains: is consumers' spatial consumption behavior important to local

[28] This is simply a consequence that individual characteristics are uncorrelated with the rain indicator. In Appendix Table C.12, we further explore triple interactions between rain and frequency of transactions with log income and with log age. As in the main table, higher income makes transactions relatively more local in high frequency sectors; additionally, we find evidence that this effect is stronger during rainy days, i.e., the triple interaction with log income is negative.

[29] In Appendix Table C.13, we replicate the analysis in this section estimating a negative binomial model, again with very similar results.

economic activity equilibrium? One way to answer this question is to study the impact of differences in local population size on local sectoral employment, as a function of demand-related sector characteristics. In this section, we present a highly stylized model that will help frame the discussion for the empirical analysis that follows. Our purpose here is not to provide a fully-estimable framework, but to highlight the key mechanisms that can drive heterogeneous employment responses to population increases. Hence, the model needs to abstract from the specifics of different sectors and focus on the moderating impact of a single sector characteristic, which we have called durability/storability.

Our framework is essentially the mirror–image of a monocentric city model: all agents live in a single place, and they choose where to shop. Since they want to consume at a constant rate over time, but travel is costly, an inventory problem emerges. We generalize ideas present in Oi (1992) – but dating back at least to Baumol (1952) – to a setting where consumers with heterogeneous travel costs choose 1) how far to travel for their purchases, 2) how frequently to do so, and 3) the purchase size per trip. A storage cost g will regulate the shape of the spatial distribution of demand for a given schedule describing the price as a function of distance. Profit–maximizing producers using a fixed factor (land) and labor will shape the spatial distribution of supply, again for a given price function. In equilibrium, a price function makes demand and supply identical point-wise and determines the marginal plot of land used.

5.1 Producers

There is one sector with productivity A. Producers operate in perfect competition and are potentially active in any location $j \in [0, +\infty)$. Each location j is endowed with a fixed amount of land $\bar{D}$. A firm located in j uses land and labor $L(j)$ to produce goods:

$$Q(j) = A\bar{D}^{1-\beta} L(j)^{\beta} \tag{6}$$

Firms located in j will choose labor to maximize profits:

$$\pi(j) = p(j)\, A\bar{D}^{1-\beta} L(j)^{\beta} - wL(j) - R(j)\, \bar{D}$$

where $p(j)$ and $R(j)$ are the output price and the rental rate of land at j; w is the wage, which we assume fixed and determined in an outside sector. The optimal quantity of labor is given by

$$L(j) = \bar{D} \left(\frac{A_i \beta}{w} \right)^{1/(1-\beta)} p(j)^{1/(1-\beta)} \tag{7}$$

and output as a function of the price, that is, the supply curve at j, is given by

$$Q(j) = A^{1/(1-\beta)} \bar{D} \left(\frac{\beta}{w} \right)^{\beta/(1-\beta)} p(j)^{\beta/(1-\beta)} \tag{8}$$

Absentee landlords will rent their land to the highest price. Under free entry, positive profits in a location will bid up these land prices; in equilibrium, the price of land is

$$R(j) = \bar{R} \cdot p(j)^{1/(1-\beta)} \tag{9}$$

and profits are zero everywhere.[30]

5.2 Consumers

A measure N of consumers is heterogeneous in t, an increasing index of individual travel costs. All consumers (exogenously) live in location 0. Each agent wants to consume a fixed quantity $\bar{q}$ of the good over one unit of time: in particular, in a fraction di of the unit time, the consumer eats a fixed quantity $1 \times di$ of the good. The assumption of a fixed quantity consumed will allow us to emphasize the role of the price function $p(j)$ in allocating consumers across distances.

A consumer with travel cost t wants to minimize the total cost of consuming $\bar{q}$ units of the good per unit of time, $c(t)$.[31] In particular, the consumer takes as given $p(j)$ and solves:

$$c(t) = \min_{j,z} p(j)\,\bar{q} + \kappa(j;t)\,\frac{\bar{q}}{z} + g\frac{z}{2} \tag{10}$$

In this expression, $\kappa(j;t)$ is the cost per trip for a consumer t traveling to j, with $\kappa(j;t) \geq 0$, $\kappa_j > 0$, $\kappa_t > 0$, $\kappa_{jt} > 0$. The consumer chooses the distance traveled j and how much to buy every trip (the "batch size") z. A batch size z implies average inventory holdings of $z/2$, and hence total inventory costs of $gz/2$; a batch size z also implies $\bar{q}/z$ trips per period (ignoring integer constraints) and hence $\kappa(j;t)\,\bar{q}/z$ travel costs. For a given distance traveled j, the consumer balances inventory costs (increasing in z) and travel costs (decreasing in z). This is just the classic trade-off in optimal inventory models, and delivers an optimal batch size of

$$z(j;t) = \left(\frac{2\bar{q}\kappa(j;t)}{g}\right)^{1/2} \tag{11}$$

Consumers will buy more per trip when the travel costs are high (to economize on the number of trips) or when the storage costs are low (to take advantage of the durability of the good). Using this expression in (10), the cost minimization problem becomes

$$c(t) = \min_{j} p(j)\,\bar{q} + \kappa(j;t)^{1/2}\,(2\bar{q}g)^{1/2} \tag{12}$$

where $\kappa(j;t)^{1/2}\,(2qg)^{1/2}$ is the total travel and storage cost associated to a given distance j under the

[30] In this equation, $\bar{R} \equiv w^{-\beta/(1-\beta)}\,(1-\beta)\,\beta^{\beta/(1-\beta)} A^{1/(1-\beta)}$ is a constant independent of our parameters of interest.

[31] We can think of this as part of a more general problem where consumers have a (large enough) income spent on this sector and on an outside good left out of the analysis.

optimal batch policy. The optimal distance traveled will satisfy

$$\frac{1}{2}\kappa\left(j;t\right)^{-1/2}\kappa_j\left(j;t\right)\left(2\bar{q}g\right)^{1/2}=-p'\left(j\right)\bar{q} \tag{13}$$

A marginally longer trip makes consumers save $-p'(j)$ per unit purchased;[32] however, consumers pay more in travel costs and inventory costs (since they optimally buy larger batches).

Note that a higher storage cost g lowers the optimal batch size (from (11)), increases the required number of trips $\bar{q}/z$, and so raises the marginal cost of distance in (13): naturally, consumers are willing to travel less for high storage cost/low durability items than for low storage costs/high durability ones, everything else equal.

5.3 Equilibrium price

We now solve for the equilibrium price function and the associated spatial distribution of production.[33] Since the price function will be the solution to a second order differential equation, we make three assumptions that allow for analytic results (and point out where they are helpful): we specialize the travel cost function to $\kappa(j;t) = (jt)^2$, we assume that $t \sim Uniform[1,2]$, and set $\beta = 1/2$ in the production function.

Our first assumption lets us write the first order condition for a consumer t as,

$$t=-p'\left(j\right)\left(\frac{\bar{q}}{2g}\right)^{1/2} \tag{14}$$

This equation explicitly assigns to each distance j a unique consumer type $t(j)$ that chooses to travel there. For a given (monotonically decreasing and convex) price function $p(j)$, this equation also implicitly assigns a unique location $j(t)$ to any consumer type t, with $j'(t) < 0$. The economy is effectively solving an assignment problem where the equilibrium matching function $j(t)$ determines a distance for any consumer type.

In equilibrium, demand for goods is equal to the supply of goods in any location where consumers choose to travel. Using (14) and the distributional assumptions on t,

$$\Pr\left\{j<\bar{J}\right\}=\Pr\left\{t\geq -p'\left(j\right)\left(\frac{\bar{q}}{2g}\right)^{1/2}\right\}=2+p'\left(j\right)\left(\frac{\bar{q}}{2g}\right)^{1/2}$$

Hence, the density of those who travel to j is given by

$$n\left(j\right)\equiv p''\left(j\right)\left(\frac{\bar{q}}{2g}\right)^{1/2} \tag{15}$$

Since each of these agents demands $\bar{q}$ units, and the population measure is N, quantity demanded at j is

[32] In equilibrium, price will in fact decrease over j.

[33] All necessary derivations in this subsection and the next are reported in Appendix B, p. 52.

$N\bar{q} \cdot n(j)$. Equating this demand to supply (8), equilibrium in good i's market in location j requires

$$p''(j) = \alpha^2 p(j), \text{ with } \alpha \equiv \alpha_0 \frac{g^{1/4}}{\bar{q}^{3/4} N^{1/2}}, \tag{16}$$

where α_0 depends on parameters.[34] This condition must hold for any location j where i is produced. Hence, (16) is a second order differential equation in the price function $p(j)$.

Definition 1 *An equilibrium is a price function $p(j)$ and a cutoff allocation of land $j_{\max}$ such that a) producers maximize profits, b) the marginal land owner obtains zero rents, c) consumers optimally choose distance, and d) demand and supply of goods are the same at every point j.*

The generic solution to (16) is

$$p(j) = c_1 \exp\{\alpha j\} + c_2 \exp\{-\alpha j\} \tag{17}$$

Using (14), the distance $j(t)$ chosen by any agent t implicitly solves

$$\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{t}{\alpha} = c_2 \exp\{-\alpha j\} - c_1 \exp\{\alpha j\} \tag{18}$$

We pin down the constants of integration using implications of our conjectured land allocation. In particular, since $j(t)$ is decreasing, the person with the lowest travel cost $(t = 1)$ will travel the maximum distance, $j_{\max} \equiv j(1)$. We use this fact in (18). We also impose in (17) that, at $j = j_{\max}$, the price of the product will have to be zero (otherwise, rents $R(j)$ would be positive, and some firms would have an incentive to enter slightly farther, paying zero rent). These steps give us two equations in the two unknown constants, and deliver

$$p(j) = \frac{1}{2^{1/2}\alpha_0} \cdot \bar{q}^{1/4} g^{1/4} N^{1/2} \cdot \left[\exp\{\alpha(j_{\max} - j)\} - \exp\{-\alpha(j_{\max} - j)\}\right] \tag{19}$$

The value of $j_{\max}$ is uniquely pinned down by imposing in (18) that the person with the highest travel cost travels as little as possible. The value $j_{\max}$ is an implicit function of α. However, we show that

$$j_{\max}(\alpha) : \frac{\alpha}{j_{\max}} \frac{dj_{\max}}{d\alpha} = -1 \tag{20}$$

Note that this implies that $\alpha \cdot j_{\max}(\alpha)$ is constant with α. Also, $p'(j) < 0$ for $j \in [0, j_{\max}]$.

5.4 Aggregate Implications

We now explore the equilibrium interactions between population size and storage costs. Note that output, employment and demand density in a location, (7), (8) and (15), are all functions of the price at the same

[34] In particular, $\alpha_0 \equiv 2^{-1/4} A \left(\bar{D}/w\right)^{1/2}$. Our third assumption of $\beta = 1/2$ allows us to write this differential equation as linear in $p(j)$.

location, for which we have an expression from (19)-(20).

We start from gravity. Expenditure at a given place j is given by

$$X(j) \equiv Nn(j) \cdot \bar{q}p(j) = \bar{x} \cdot p(j)^2$$

where the last equality uses (15) and (16).[35] Since the price is decreasing in j, expenditure decreases with distance and gravity holds. Similar to continuous types–labor market models of assignment (e.g. Sattinger, 1979; Costinot and Vogel, 2010; Monte, 2011), the model delivers a non-linear value of different locations that depends on the complementarity between j and t, and on the distribution of t.[36]

Not only does gravity hold, but it is steeper in sectors with high storage costs. Consider the simple slope $[X(j_{\max}) - X(0)]/j_{\max} = -X(0)/j_{\max}$. When g is higher, the marginal cost of distance for all consumers grows and the willingness to take long trips shrinks. This implies that the marginal plot of land is no longer viable, and $j_{\max}$ falls.[37] Since output is still fixed at Nq but there is less land, more demand must be concentrated at shorter distances, in particular at $j = 0$, so that $p(0)$ grows.[38] The average slope of the expenditure is then higher when g grows.[39]

Higher storage costs also induce each consumer to buy more frequently. To see this, it is sufficient to consider the expression for the optimal batch size z in (11): as g grows, a consumer will reduce z for any distance traveled. She will also reduce the distance traveled $j(t)$, since the marginal cost of distance is higher. Hence, z unambiguously drops, and the frequency q/z unambiguously rises.[40] These arguments show that the model generates a negative relation between frequency and gravity as in Figure (3).

How does the spatial distribution of output and employment respond to a bigger population? We start by noting that if population N increases, total output $N\bar{q}$ increases in the same proportion, by assumption of fixed $\bar{q}$. The expansion in demand and production requires more inputs. The total amount of land used $j_{\max}(\alpha)$ also increases, but more slowly than 1–for–1 (in particular, with elasticity 1/2). Hence, in response to an increase in market size, output density $N\bar{q}/j_{\max}$ must grow: intuitively, since consumer travel is expensive, firms try to increase output per unit of land, not just total land used.

Given that the marginal plot of land used is farther away, the growth in population increases the

[35] In this expression, $\bar{x} = 2^{-1/2}\alpha_0^2$ is a function of parameters not involving g or N.

[36] A related setting is studied in Karádi and Koren (2017): they focus on the general equilibrium implications of sectoral location choice (our model is partial equilibrium in that we only study equilibrium in one product market), where, however, the impact of distance is modeled as a classic iceberg decay; hence, expenditure declines log-linearly with distance. In this sense, this framework is related to demand–side non-linear pricing models in international trade: for example, Fieler (2011) or Fajgelbaum, Grossman and Helpman (2011). See Sattinger (1993) for a discussion of the role of linear vs. non-linear pricing on income distribution in assignment models. Our framework also generates a non-constant distance elasticity; for an explanation of why the distance elasticity in international trade is constant and close to -1, see Chaney (2018).

[37] Recall that $j_{\max}$ decreases with α, and α increases with g.

[38] Using (19), it's easy to see that $p(0) \propto g^{1/4}[\exp\{\alpha j_{\max}\} - \exp\{-\alpha j_{\max}\}]$, which increases with g since $\alpha j_{\max}$ is constant.

[39] Since the equilibrium expenditure function has a varying elasticity over j, it is difficult to prove that the slope becomes steeper at every j. However, Lemma 2 shows that over $j \in [0, j_{\max}]$ 1) the unweighted average slope $X'(j)$, 2) the average slope $X'(j)$ weighted by the number of agents $n(j)$, and 3) the average slope of $X'(j)$ weighted by the expenditure at j, $X(j)$, all become more negative as g grows.

[40] See Lemma 3 on p. 54 for a proof.

average distance of output to consumers:

$$\frac{\int_0^{j_{\max}} jQ\left(j\right)dj}{Nq} \equiv \bar{d}\cdot\frac{N^{1/2}}{g^{1/4}}, \tag{21}$$

where $\bar{d}$ depend on parameters others than g and N.[41] However, travel is more expensive for high storage-cost sectors: hence, the average distance increases less when g is higher.

Given that land per location is fixed, the increase in output density must be generated via changes in employment density. Using (7), total employment in the sector is given by

$$L_{eq} = \int_0^{j_{\max}} L\left(j\right)dj = \frac{\bar{L}}{Aw^{1/2}}\cdot g^{1/4}N^{3/2} \tag{22}$$

where $\bar{L}$ depends on parameters. This expression is intuitive. As population increases, employment must grow more than 1–for–1 to limit the increase in $j_{\max}$ and economize on travel costs. This increase in employment is larger when storage is more expensive, that is when g is high, because the need to economize on travel is stronger. Armed with these intuitions, we turn to an empirical exploration of consumer demand on local outcomes.

6 Consumers Demand and Local Equilibrium Outcomes

We have provided above an intuitive reason for why higher storage costs make demand more "local": consumers would rather buy smaller batches with more frequent trips; since travel is expensive, however, consumers will optimally choose to take shorter trips. Hence, in response to a larger local population, we might expect to see local employment in high storage-cost sectors grow relatively faster than employment in sectors with low storage costs: firms are trying to economize on consumer travel time by limiting the amount of distant land used, and to do so they substitute land with labor. In addition, if savings in travel time are at the root of this behavior, we might expect to see employment growth driven by a higher density of stores (that is, a reduced average distance between consumers and stores), rather than more employees per store.

In this section, we study the impact of differences in population on county–sector employment as a function of the sector's average number of transactions, our simple proxy for storage costs. We emphasize that we see this exercise as guided by, but not strictly testing, our simple model. We use our stylized framework as an organizational device that lacks many of the general equilibrium links, but that can be useful in organizing our thinking around the consequences of consumer mobility. We see the forces we describe playing out in a long run equilibrium, after the entry-exit margin of new establishments has been allowed to adjust. Hence, our main empirical examination will leverage cross-sectional differences across space and sectors.

To explore the heterogeneous response of local sectoral outcomes to local population we estimate

[41] See Lemmas 4 and 5 on pp. 55 and following for a proof of eq. (21) and (22).

regressions of the form

$$\ln y_{sct} = \alpha + \beta \ln pop_{ct} + \gamma \ln freq_s \times \ln pop_{ct} + \eta_0 \ln i_{ct} + \eta_1 \ln size_c + FE + \varepsilon_{sct} \tag{23}$$

In this regression, s indexes MCC sectors, c indexes counties, and t denotes calendar year ($t = 2007$ and 1998). The regressor $\ln y_{sct}$ may assume three values. We first use log employment in s, c, t: to construct it, we have started with data in the relevant years from County Business Patterns, and have developed a correspondence between NAICS 6 digits and MCC codes. Always using County Business Pattern data, we also explore the response of the number of local establishments $\ln y_{sct} = \ln n_{sct}$, and employees per establishment $\ln y_{sct} = \ln (emp_{sct}/n_{sct})$; the regressor $\ln freq_s$ is the log average frequency of transactions in sector s across all accounts in the credit card data; $\ln pop_{ct}$ and $\ln i_{ct}$ are the county log population and average personal income per capita, from the County Economic Profile of the Bureau of Economic Analysis; $\ln size_c$ is the county land area; FE is a set of fixed effects, varying across regressions (sector fixed effects are always included and absorb the regressor $\ln freq_s$ in levels); and ε_{sct} is a stochastic unobserved term.

Table 7 reports OLS estimates of Equation (23) jointly for 1998 and 2007. In the first two columns, the dependent variable is county-sector log employment; columns (3) and (4) use total number of establishments, and columns (5) and (6) use the average number of employees per establishment. Each group of two columns differ by the set of fixed effects used: columns (1), (3) and (5) use sector fixed effects and year fixed effects; columns (2), (4) and (6) allow for county-year and sector-year dummies, i.e., control for heterogeneous time trends within sectors and within counties.

As one might expect, and as our previous intuition suggests, population enters positively. Moreover, conditional on observables, sectoral employment grows less with population in sectors that are transacted more frequently. Columns (3)-(6) indicate that this slower employment growth occurs for about two thirds via smaller stores, and about one third via fewer stores in the county. Our simple model would suggest, however, that as population grows, sectors whose demand is more spatially concentrated should see relatively faster local employment growth. This disconnect is not necessarily surprising, since our OLS regression does not address the presence of possible endogeneity. In the remainder of this section, we explore theoretically and empirically these factors, and implement an alternative identification strategy.

6.1 Limits of OLS and Identification

Two natural features of economic activity that are absent from our stylized model are the fact that population and productivity may be related in equilibrium, and the presence of local labor market equilibrium links. We start by extending our simple model to account for these two features. Let's assume that equilibrium population and productivity are related via $N = \bar{N} A^{\zeta}$, where $\bar{N}$ is an exogenous component and $\zeta \geq 0$ captures in a reduced-form way the positive relation between local productivity and population. This relation may come via agglomeration externalities or because high productivity reduces the local consumer price index. Also, assume that labor supply has a positive but finite supply elasticity $\varepsilon > 0$ with respect to wages w, i.e. $L^s = N w^{\varepsilon}$. Equating labor demand L^d (now described by eq. (22)) to labor

Table 7: **Local outcome responses (OLS estimates)**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log land area	0.125***		0.122***		0.003	
	(0.010)		(0.007)		(0.006)	
Log income per capita	1.112***		0.934***		0.178***	
	(0.035)		(0.025)		(0.020)	
Log population	1.216***		0.918***		0.298***	
	(0.007)		(0.005)		(0.004)	
Log population × log frequency	-0.064***	-0.070***	-0.021***	-0.021***	-0.043***	-0.050***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Sector Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	Yes	No	Yes	No	Yes	No
Sector-Year Fixed Effects	No	Yes	No	Yes	No	Yes
County-Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-square	0.84	0.88	0.89	0.92	0.52	0.58
N	121,336	121,333	121,336	121,333	121,336	121,333

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

supply with these two extensions,

$$L^d = \bar{L} \cdot g^{1/4} \bar{N}^{3/2} A^{3\zeta/2-1} w^{-1/2} = \bar{N} A^{\zeta} w^{\varepsilon} \tag{24}$$

This equation determines an equilibrium wage,[42] which we can then substitute in eq. (22) to obtain

$$L_{eq} = \hat{L} \cdot g^{\frac{1}{2}\frac{\varepsilon}{(2\varepsilon+1)}} \bar{N}^{\frac{3\varepsilon+1}{2\varepsilon+1}} A^{\frac{(3\varepsilon+1)\zeta-2\varepsilon}{2\varepsilon+1}} \tag{25}$$

where $\hat{L}$ is again a function of parameters. An increase in productivity A will increase population, since $N = \bar{N} A^{\zeta}$; when this relation is not too strong (in particular, $\zeta < \frac{2\varepsilon}{3\varepsilon+1}$), an increase in productivity will, quite naturally, reduce equilibrium employment. This intuition will hold as long as labor supply is not perfectly rigid (i.e., $\varepsilon > 0$). Moreover,

$$\frac{\partial L_{eq}}{\partial g \partial A} < 0 \tag{26}$$

In other words, an unobserved productivity increase could generate both higher population (since $N = \bar{N} A^{\zeta}$) and employment that grows relatively less (or falls relatively more) in high g sectors than in low g sectors, since high g sectors have the largest local employment to begin with.

With this intuition in hand, we inspect the source of the negative sign on the interaction between pop-

[42] In particular, $w = \left(\bar{L} \cdot g^{1/4} \bar{N}^{1/2} A^{\zeta/2-1}\right)^{2/(2\varepsilon+1)}$

ulation and frequency in Table 7 by examining the behavior of employment across counties with different initial characteristics. In particular, we group counties in 5 bins according to their initial population density in 1998: if county population and density are related to productivity (e.g. Combes et al., 2012), we might expect this coefficient to vary with the set of counties included in the regression.

We then re-estimate columns (2), (4) and (6) in Table 7 progressively for the set of counties up to and including the q^{th} quintile, with $q = 1, ..., 5$. Our findings conform to this intuition. The left panel of figure 4 plots the coefficient γ_q on the interaction between population and frequency including progressively more quintiles. The panel shows that in low density counties, a marginal increase in population is associated with larger employment in high frequency compared to low frequency sectors. This larger employment occurs via an increase both in the number of establishments and in the average establishment size. Hence, we find support for the implications of our stylized model on the cross-sectoral behavior of employment in this subset of counties. As we include progressively denser counties, however, the effect becomes smaller and then it reverses. The right panel of Figure 4 indicates the marginal effects on the two dimensions of number of establishments and establishment size: the reversal occurs via both margins.

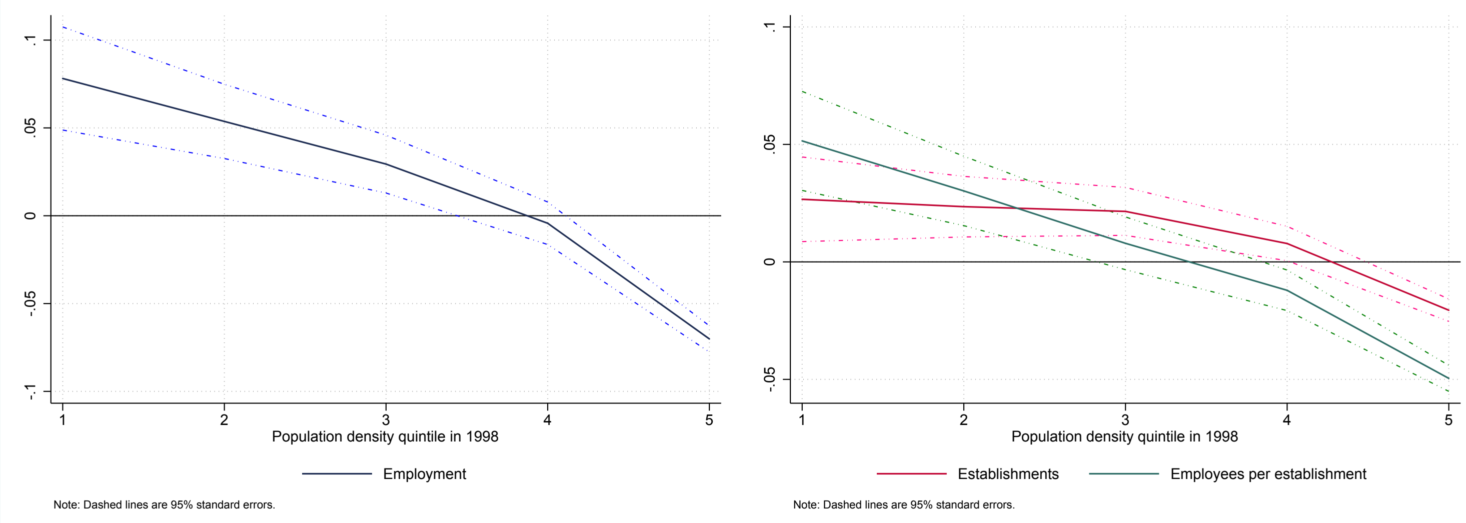

Figure 4: Slope of interaction of population and frequency (OLS)

The analysis of potential sources of endogeneity also suggests a potential alternative strategy. Eq. (25) indicates that if we were able to induce exogenous variation in $\bar{N}$, we should expect a positive cross partial between population and g, since:

$$\frac{\partial L_{eq}}{\partial g \partial \bar{N}} > 0 \tag{27}$$

Note that this cross-partial derivative should be positive even in presence of labor market equilibrium feedbacks: hence, the intuition of our simple model carries through in a more general data-generating process where equilibrium employment responds to local labor supply and wages.[43]

[43] In a later section, we further show the differential behavior of employment when we study the consequences of increases in non-working population only, and increases in population "corrected" for the role of commuting flows.

To address these concerns, we choose an instrumental variable strategy that exploits the underlying geological composition of a county to induce variation in $\bar{N}$. In particular, we borrow an intuition developed in Burchfield, Overman, Puga, and Turner (2006) and Duranton and Turner (2017) and compute the fraction of land of a county laying over consolidated and semiconsolidated aquifers.[44] An aquifer is an underground layer of water–bearing rock, and the presence of different types of aquifers induces quasi-random variation in the population residing over the territory of a county. Hence, we instrument county population with those fractions, and the interaction of population and frequency of transactions with the interaction of the same percentages with the sector-level frequency of transactions.

Since the instrument is time–invariant, it induces quasi-random variation in population across counties over space but not within counties over time: hence, it is consistent with our intention of leveraging cross–sectional differences in a long run equilibrium.

Our instrumental variable is valid if aquifers affect employment only through population. For fixed land area, however, the presence of aquifers increases both population and density by construction.[45] Therefore, a possible threat to our identification strategy emerges if density has a differential effect on sectoral employment which is independent from population. For example, one might imagine that people with preferences for high frequency amenities select into cities which tend to be denser and older; in this case, an exogenous increase in density might see an employment change that arises from preferences rather than from purchasing behavior of consumers. We will assuage these concerns by examining how our instrumental strategy performs as we again start from the least dense and progressively include denser counties in our analysis; in robustness checks, we will also allow directly for independent heterogeneous effects of average county income or of density in a regression analysis.

In the next section, we describe the results of implementing this approach.

6.2 Instrumental Variables Estimates

We are interested in identifying the causal effect of population on the differential employment growth across sectors. We instrument county population with information on the underlying geological composition of the county. More precisely, we follow Duranton and Turner (2017), compute the fraction of land laying over consolidated and semiconsolidated aquifers and use those percentages as instruments for population; moreover, we instrument the interaction of population and frequency of transactions with the interaction of the same percentages with frequency. In sum, we instrument 2 endogenous variables with 4 instruments. Our main results are reported in Tables 8-10, which estimate eq. (23) using our instrumental variable approach. In these tables, standard errors are clustered at the county level.

The strategy results in a good first stage across all specifications. Across all tables (see for example the bottom two rows of Table 8) we report the Cragg-Donald (CD) Wald F-statistic for the strength of the first stage identification. This statistic can be evaluated against the Stock and Yogo (2005) critical

[44] This geological information comes from the United States Geological Service, Principal Aquifers of the 48 Conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands. We use standard geoprocessing software to compute the county composition.

[45] Indeed, Duranton and Turner (2017) use instruments based on this data as an exogenous shifter for density of population.

values for 2 endogenous variable with 4 instruments.[46]

Column (1) of Table 8 shows that, after controlling for endogeneity, the sign on the interaction coefficient reverses: in 2007, when the population of a county is larger because of underlying geological reasons, the effect on employment is larger in a sector with high storage costs than in a sector with low storage costs. Moving from the minimum to the maximum average frequency changes the growth in employment by 2.5 percentage points for a 10% increase in population. This is equivalent to $100{\cdot}2.5/7.79 =$ 32% of the estimated baseline increase in employment for the lowest frequency sector.[47] Column (2) and (3) show the same regression run in 1998 and for the stacked sample of two cross-sections with year fixed effects. Obviously, time trends may be operating differentially for different areas and sectors, and this may affect our estimates in the stacked regression. In column (4) and (5), we allow for heterogeneous time trends across sectors (both columns), and across U.S. states (column (4)) or commuting zones (column (5)). As we include more detailed geography fixed effects, the cross-sectional variation that we exploit becomes absorbed to a larger degree. Nonetheless, the coefficient on the interaction term stays positive and significant. In the most restrictive specification, where we only exploit geological differences between counties within the same commuting zone, moving from the smallest to the largest frequency changes the employment response by 1.72 percentage points per 10% increase in population, or 14% of the baseline impact of population. In all cases, the CD statistics are above the Stock-Yogo (2005) 5% significance critical values for a 5% maximal bias of the 2SLS estimate relative to the OLS (critical value 11.04).[48]

The regressions shown so far indicate that if storage costs are reasonably proxied by the observed frequency of transactions, consumers' spatial choices have relevant consequences on economic outcomes: our model suggests that, in response to larger population, all sectors want their output to grow; however, the desire to economize on distance is stronger (and hence the substitution of land with employment growth is greater) in sectors where storage costs are high. To investigate how this equilibrium impact comes about, we ask next how employment is increasing in the county. In practice, an increase in local sectoral employment may be generated entirely at the intensive margin, i.e., via more employees per store. If time savings are important, however, we might expect that demand in high storage cost sectors is served by a higher density of stores, i.e., via a lower average distance between consumers and stores. Our model does not have a specific prediction about this mechanism, but suggests that the average distance between consumers and output should be smaller (and grow less with population) in high–g sectors (Equation (21)); in other words, the county density of stores, an inverse proxy of the distance between consumers and establishments, might be growing faster in high g than in low g sectors.

Table 9 and 10 show that indeed, the geographical concentration of stores grows relatively faster with population in high storage-cost relative to low storage-cost sectors. In particular, Table 9 replicates Table 8 but uses the log number of establishments in a given county-sector-time as a dependent variable.

[46] We also report the Kleibergen-Paap Wald F statistic that accounts for clustering in the standard errors. Critical values for this statistic have not been tabulated in the literature.

[47] The coefficient β on $\ln pop_{ct}$ is influenced, in general, by all other general equilibrium effects coming through a larger market. Here and below, we compare the impact of the range in log frequency to the baseline to give a broad sense of the magnitude, but such an effect should be interpreted with caution.

[48] Our robustness section discusses alternative ways of exploring the presence of weak instruments.

Table 8: **Local employment and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			county-sector log employment		
Sample years :	07	98	98,07	98,07	98,07
Log population	0.773***	0.905***	0.835***	0.820***	1.225***
	(0.095)	(0.071)	(0.080)	(0.123)	(0.084)
Log population × log frequency	0.142***	0.104**	0.124***	0.114***	0.098***
	(0.048)	(0.042)	(0.043)	(0.042)	(0.038)
Log income per capita	1.614***	1.601***	1.622***	1.766***	1.016***
	(0.140)	(0.136)	(0.134)	(0.295)	(0.202)
Log land area	0.104***	0.136***	0.120***	0.183***	-0.024
	(0.012)	(0.011)	(0.011)	(0.034)	(0.043)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.82	0.84	0.83	0.83	0.85
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	205.05	223.19	435.95	233.1	346.08
Kleibergen-Paap rk Wald F statistic	23.42	21.73	24.93	5.68	6.15

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Estimates on the interaction coefficient are now two to three times larger and strongly significant. These results are consistent with a situation where, in response to a common increase in population, the increase in demand is more geographically concentrated for high storage costs goods and services, where people desire frequent transactions and shorter trips; the supply side then responds by increasing employment via a relatively denser presence of stores. The most conservative estimates imply that the highest storage-cost sector has 2.54 percentage points larger number of establishments relative to the lowest storage-cost sectors, when population increases 10% (about 29% of the baseline impact of population). Table 10 shows that, if anything, establishments become relatively smaller on average: using estimates in column (5), the size of the average store in the highest storage cost sector grows 0.82 percentage points slower relative to the lowest storage cost sector, when population grows 10% (about -21% of the baseline impact of the population).

To understand more about how the instrument works and how it flips the sign of the interaction, we re-estimate our IV regressions starting from the quintile of counties with lowest density and then progressively including all the others (mimicking the construction of Figure 4 above). The results are reported in Figure 5. The instrument operates by raising the estimated net effect of population times

Table 9: **Number of establishments and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.492***	0.606***	0.546***	0.560***	0.814***
	(0.079)	(0.057)	(0.066)	(0.092)	(0.059)
Log population × log frequency	0.180***	0.152***	0.167***	0.155***	0.145***
	(0.040)	(0.036)	(0.037)	(0.035)	(0.032)
Log income per capita	1.440***	1.372***	1.418***	1.435***	0.939***
	(0.117)	(0.101)	(0.106)	(0.215)	(0.133)
Log land area	0.100***	0.135***	0.118***	0.160***	0.035
	(0.011)	(0.008)	(0.009)	(0.026)	(0.029)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.86	0.88	0.87	0.87	0.90
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	205.05	223.19	435.95	233.1	346.08
Kleibergen-Paap rk Wald F statistic	23.42	21.73	24.93	5.68	6.15

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

frequency throughout the range of densities.[49] The sign of the interaction coefficient using the full sample (at quintile 5 in the left panel of Figure 5) switches from negative to positive because the instrumental variable approach estimates a stronger positive effect on the extensive margin of the number of stores, and a weaker negative effect on the store size.

We now return to a consideration of the threats to our identification strategy. For given land area, an increase in population is also, by construction, an increase in density. Hence, in principle we may be comparing a large (and high density) place to a small (and low density) place. It could be the case that density has an independent differential effect across sectors on top of population. One might imagine that aquifers are associated with older cities which are now denser, scarcer in space and where both residential and commercial rents are high; it might be the case that the type of people that sort into these places have a stronger preference for high frequency sectors with respect to people that demand more space and possibly sort into newer and maybe less dense cities. If this is true, places with higher density will attract

[49] The coefficients on the interaction terms reported in these figures are estimated using the specification in column 6 of Tables 8-10, i.e., including at most commuting zone-year and sector-year fixed effects. Strictly speaking, then, these are not directly comparable with Figure 4 above, which report the OLS estimates using county-year and sector-year fixed effects. However, if we replicate that Figure 4 using the OLS version of column 6 of Tables 8-10, differences are unnoticeable.

Table 10: **Number of employees per establishment and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.281***	0.299***	0.288***	0.260***	0.411***
	(0.041)	(0.036)	(0.035)	(0.054)	(0.047)
Log population × log frequency	-0.038	-0.048**	-0.043**	-0.040*	-0.047**
	(0.025)	(0.025)	(0.022)	(0.022)	(0.021)
Log land area	0.004	0.002	0.002	0.023	-0.058**
	(0.006)	(0.006)	(0.006)	(0.015)	(0.026)
Log income per capita	0.174***	0.229***	0.204***	0.331***	0.077
	(0.059)	(0.064)	(0.058)	(0.127)	(0.117)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.51	0.53	0.52	0.52	0.54
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	205.05	223.19	435.95	233.1	346.08
Kleibergen-Paap rk Wald F statistic	23.42	21.73	24.93	5.68	6.15

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

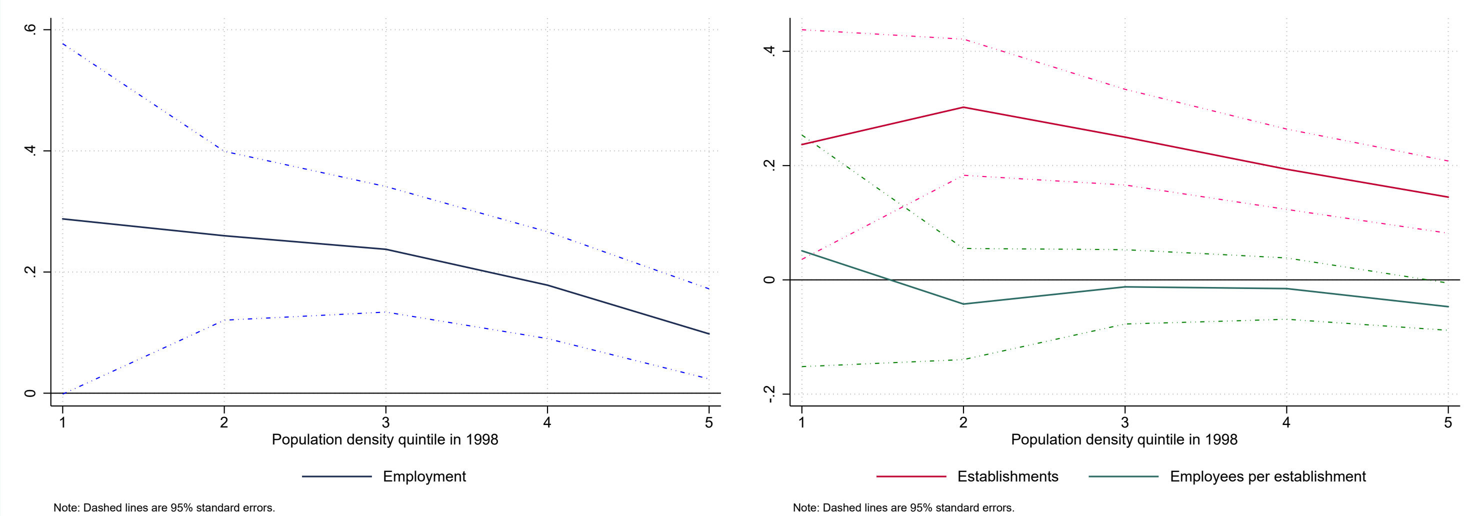

Figure 5: Slope of interaction of population and frequency (IV)

high-frequency types of economic activity relatively more.

A series of considerations indicate that these concerns, while obviously valid, may be of more limited practical importance. First, the battery of regressions in Table 8-10 include a set of increasingly narrow geographic fixed effects: in the strictest specification, we are comparing counties within the same commuting zone, rather than big and dense commuting zones vs. small and sparse ones. Second, if the story we described was empirically relevant, any heterogeneity in the impact of density would let the coefficient on the interaction term to *grow* as we progressively include denser counties; however, Figure 5 shows that, if anything, the coefficient becomes smaller. Third, we can run regressions that directly control for a differential effect of income-varying preferences or of density. In Appendix C.8, we report more details about these exercises. As a summary, we first control for an interaction between income per capita and log frequency of transactions to allow for heterogeneity in the role of income on local demand. Table C.14 shows that higher income is associated to a relatively smaller employment in high frequency sectors, while the coefficient on population times frequency is little affected. In Table C.15 we then control for density, and its interaction with frequency, directly: hence we are comparing, say, a small sprawling place to a big sprawling place. The interaction between density and frequency is *negative*, that is, a denser place has relatively less employment in high frequency sectors than in low frequency ones; this negative sign is consistent with the evidence presented in Figure 5, that shows a declining slope on the interaction as denser counties are included in the sample. Importantly, the coefficient on the interaction between population and frequency stays positive and significant.

Taken together, these results paint a picture consistent with the importance of consumer mobility for local economic outcomes. In sectors with high storage costs, consumers should be more willing to trade off larger batches with frequent trips, but to do so they would choose to travel shorter distances. An implication is that, in response to larger population, firms in high storage-cost sectors face a relatively stronger incentive to increase production by economizing on land; hence, local employment should increase relatively more. We find support for these implications on employment in the data. This support appears present both in a subset of counties where endogeneity concerns may be more muted, and across all counties when an instrumental variable strategy offers a way of addressing such endogeneity. Employment responds via a denser network of local suppliers, which reduces the average distance between consumers and shops. This response is compatible with the importance of time savings for consumers in sectors more frequently transacted. We emphasize that our instrumental strategy is moving population and density together; however, a series of considerations increase our confidence that our results are not driven exclusively by an independent role of density.

We close our empirical work with further robustness and caveats to the interpretation of our results.

7 Robustness and limitations

In this section we explore some further robustness checks, and identify some possible shortcomings of our analysis.

Commuting-adjusted population. It is natural to think that commuting may affect spatial spending patterns. Workers may make some of their purchases near their workplace or along their commuting path. Moreover, households overall may become more familiar with shopping options next to the workplace locations of some of their members. Our credit card data contains no information on the workplace location of their account holder. We then try to explore this dimension by constructing a measure of commuting-adjusted population. In particular, we multiply the county population by the share of residents who work in the same county from Monte et al. (2018). Although imperfect, this measure has the advantage of capturing better the heterogeneity across counties in the degree of local expenditure arising from deeper knowledge of local environment, or joint work-shopping decisions. On the other hand, this measure has the disadvantage of introducing a stronger feedback via labor supply linkages. Our measure of "likely pure consumers below" provides a complementary point of view.

Table 11 reports the results for the total employment response. We start by noting the effect on the baseline employment, which is now around 13 to 20 log points larger in almost all specifications. These findings support our intuition that the commuting-adjusted population measure captures to a larger degree variation induced by home-workplace commuting within counties, and their related labor-supply equilibrium effects.[50] The differential response of employment as a function of frequency is again positive and larger than the corresponding baseline specifications in Table 8 by 2 to 4 log points. This increase is generated by even denser but smaller stores (Tables C.16 and C.17 in Appendix C.9) with respect to the baseline.

Likely pure consumers. A potential concern with our results is that in the data, an increase in population generates both an increase in consumers and an increase in employable workers. The increase in workers, by changing labor supply and hence wages, might affect the incentive of firms to expand; one might further imagine that the expansion is correlated with sectoral characteristics that are in turn correlated with the frequency of transactions. In this case, one could measure a positive interaction between frequency of transactions and population for reasons unrelated to consumer behavior. We have argued that within our stylized model, our identification is robust to labor supply feedbacks; however, other forces outside the model may still be operating in a way that generate the described correlation.

To further address this concern, we construct a measure of "likely pure consumers", that is, population that lives in the county but is not employed anywhere. To compute this value, we subtract a measure of county residential employment (residents of a county working inside or outside the county) from the total county population in a given year. To compute county residential employment we use information contained in a given year in bilateral commuting matrices (see Appendix C.10 for more details). This measure of likely pure consumers includes for example young and retired residents, and working-age population out of the labor force.

There are two advantages of using this measure. First, by construction, changes in likely pure consumers are less likely to affect employment via labor supply, and more via changes in demand for products. Second, since likely pure consumers do not work by construction, their consumption is less likely to be

[50] A further natural consequence of giving a greater weight to employment-related considerations is that our instrumenting strategy becomes somewhat less appropriate, which is reflected in generally lower Cragg-Donald and Kleinbergen-Paap statistics.

Table 11: **Local employment and frequency of purchase: Commuting-Adjusted Population (C.a.p.)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			county-sector log employment		
Sample years :	07	98	98,07	98,07	98,07
Log commuting-adjusted Population	0.900***	1.024***	0.958***	1.108***	1.027***
	(0.108)	(0.086)	(0.092)	(0.115)	(0.085)
Log commuting-adjusted population × log frequency	0.199**	0.143**	0.172**	0.135**	0.145**
	(0.082)	(0.071)	(0.073)	(0.067)	(0.068)
Log income per capita	1.089***	1.029***	1.071***	0.774***	0.983***
	(0.118)	(0.115)	(0.111)	(0.247)	(0.155)
Log land area	-0.099***	-0.091***	-0.094***	-0.108***	-0.086**
	(0.011)	(0.012)	(0.011)	(0.039)	(0.036)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.84	0.85	0.85	0.85	0.86
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	84.43	84.22	169.16	113.12	168.14
Kleibergen-Paap rk Wald F statistic	11.92	11.76	13.11	2.83	4.03

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

driven by possible co-location of workplace and consumption choices. The disadvantage of using this measure is that workers are consumers too, and hence the empirical analysis misses this chunk of the demand. In this sense, this exercise is complementary to our analysis of Ccommuting-adjusted population.

Table 12 reports the results for the overall employment response, a parallel to Table 8. We note first that the baseline level for likely pure consumers is between 1 to 5 log-points lower than the population level: this suggests that indeed focusing on increases in non-working population reduces the significance of equilibrium effects via labor supply. We further find that the differential response of employment as a function of frequency is still positive when we focus on likely pure consumers. Moreover, the slopes of this interaction are around 4-5 log-points larger than the corresponding specifications in Table 8 above. Comparing the response of number of establishments (Table C.18 in Appendix C.10) and employees per establishment (Table C.19 in Appendix C.10) to the baseline Tables 9 and 10, we find that most of the increase in the interaction term comes from the establishment margin: in other words, when we approximate pure consumers, the importance of the density of stores grows. These findings are consistent with the view that consumer mobility contributes to shape sector-level differences in local equilibrium outcomes in terms of employment and its composition.

Fixed costs. In Tables 8-10, we have argued that an exogenous increase in population tends to generate a demand that is more geographically concentrated for high storage-cost sectors than low storage-cost

Table 12: **Local employment and frequency of purchase: Likely pure consumers (L.p.c.)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			county-sector log employment		
Sample years :	07	98	98,07	98,07	98,07
Log likely pure consumers	0.720***	0.884***	0.799***	0.796***	1.173***
	(0.112)	(0.085)	(0.095)	(0.138)	(0.095)
Log likely pure consumers × log frequency	0.205***	0.161***	0.184***	0.159***	0.143***
	(0.061)	(0.054)	(0.055)	(0.052)	(0.048)
Log income per capita	1.965***	1.966***	1.978***	2.140***	1.562***
	(0.133)	(0.129)	(0.127)	(0.274)	(0.198)
Log land area	0.069***	0.092***	0.081***	0.179***	0.000
	(0.013)	(0.012)	(0.012)	(0.037)	(0.050)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.80	0.83	0.82	0.82	0.85
N	60,350	60,828	121,178	121,178	121,178
Cragg-Donald Wald F statistic	184.99	186.57	375.1	242.64	367.93
Kleibergen-Paap rk Wald F statistic	20.83	17.99	20.78	6.12	16.01

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ones. Heterogeneous fixed costs across industries, however, may also affect the density of establishments and hence confound our estimates.

A number of considerations support the view that these issues are not a primary concern. First, as we have noted in footnote 25, the correlation between log average frequency of transactions and log fixed costs across sectors is only 0.12, and statistically indistinguishable from zero: hence, it is not empirically true that low fixed cost sectors are those where transactions are more frequent. Second, our analysis uses sector fixed effects which absorb any factor varying at the industry level, including the true measure of fixed costs. Finally, we can explore the sensitivity of our analysis to the interaction of fixed costs with population. Similar to storage costs, direct observations of fixed costs are hard to obtain. However, a reasonable proxy is the economy-wide ratio of total employment to total establishments in a sector-year, i.e., the average establishment size. If fixed costs are high, increasing returns to scale are more important, and we should expect a higher employees-to-establishment ratio. This new interaction variable is again instrumented with the interaction between fixed costs and county geological composition.

If fixed costs are driving the results via their effect on local density, we should expect the interaction of population with log frequency to lose significance. Table C.20 in Appendix C.11 replicates the most conservative specifications in columns (5) and (6) for Tables 8-10. The coefficient on the interaction between frequency and population stays positive and of very similar magnitude. Overall, we read these results as further evidence consistent with a role of consumers' mobility on local economic outcomes.

Dependence on particular sectors or geographical areas. In Appendix C.12, we examine the extent to which our results depend on the inclusion of particular sectors or areas. In a first exercise, we re-estimate the strictest specification of Column (6) for Tables 8-10, by excluding one sector at a time. Table C.21 shows that in no case do our estimates of the interaction between log population and log frequency loses significance. In a second exercise, we examine whether our results depend on particular geographical areas. We select 5% of commuting zones at random, and again re-estimate Column (6) from Tables 8-10. We repeat this process 500 times. Figure C.4 shows that in no case do the estimates of the interaction coefficient falls below the 10% significance level.

Weak instruments. Our tables report Cragg-Donald and Kleibergen-Paap statistics for detecting weak instruments. An alternative way of assessing the threat of weak instruments is to compare 2SLS estimates to Limited Information Maximum Likelihood (LIML) ones. LIML has better small sample properties in presence of weak instruments, and large differences between 2SLS and LIML estimates may point to instrument weakness. We replicate in Appendix C.13 the main tables of section 6.2 and show that coefficients and standard errors are indeed very similar.

A further way to assess the robustness of our results is to use still plausible but different sets of instruments: in the same Appendix C.13 we also replicate the main tables of the section using all six aquifer types reported in the U.S. Geological Service data, rather than only consolidated and semi-consolidated aquifers. Our coefficients become somewhat smaller, reflecting the smaller power of the instruments; however, all patterns of significance are robust.

Selection into method of payment. It has been documented (see for example Wang and Wolman, 2016) that transactions of smaller dollar size tend to be executed with cash, rather than with other means. Unfortunately, our data does not allow us to control for this choice. In unreported results, we find that the average transaction value increases slightly with distance, controlling for consumer characteristics; hence, short trips are less likely to be reported in our data. On the one hand, this selection will make gravity appear less important than it actually is, since we are removing expenditure that occurs close-by; this effect will, in fact, be stronger in sectors where the average distance traveled is shorter, i.e., in sectors with a high frequency of transactions. Via this first channel, the relation between gravity and frequency documented in Figure 3 should be steeper than we measure. On the other hand, this selection will also remove more of the short trips (which are of higher frequency) than the longer trips (which are of low frequency). Via this second channel, the relation should be flatter than we measure. The fact that these two forces tend to compensate each other makes it hard to offer clear predictions on the net effect of these unobservable choices, and hence our results should be interpreted with this limitation in mind.

Trip chaining. It is natural to think that one way in which consumers optimize their shopping behavior is to make a number of possibly unrelated purchases on a single trip to a commercial area. For example, Shoag and Veuger (2017) document positive externalities of "big-box" stores on neighboring businesses via the increased local foot traffic. Our data is unfortunately too coarse to speak to that aspect:

out of all account-transaction dates in our data, only 25% have more than one transaction per day, and less than 1% have at least 5 transactions; of the cases in which there is more than one transaction, 80% are multiple transactions occurring in the same broad sector. How would our results be impacted if this was the predominant behavior of consumers? Suppose that consumers always travel to one mall and buy food every trip, but apparel every four trips. In that case, we would expect to see no relation between gravity and the frequency of transactions, since the frequency of purchase differs, while the distance stays constant. More importantly, the frequency of transactions in the credit card data would less likely predict heterogeneity in the impact of population on density. The fact that we see at least some impact is indicative that trip chaining is not the only relevant feature of the data.[51] As above, however, we emphasize that our conclusions should be considered accounting for this limitation.

8 Conclusion

Using detailed geographical information from more than 1.7 million credit card transactions by individual consumers, we document several stylized facts regarding the geography of consumption. We find considerable heterogeneity across industries in the overall impact of distance and in the importance of extensive margins. We identify a new sector characteristic that contributes to determine consumer spatial behavior – the storability/durability of the product or service. Differences in gravity across industries are correlated with the frequency of transactions, a proxy for storability/durability. This relationship suggests that consumers actively choose the spatial margin of their purchases considering the storability/durability of a sector's output, and an analysis of individual–level behavior further supports this view: sectors which are more frequently purchased tend to be transacted more locally by richer individuals, and their spatial distribution of transactions tends to be less sensitive to rain episodes.

Finally, we present evidence consistent with consumer mobility having implications for local equilibrium economic outcomes like employment, store density and store size. In response to larger population, sectors with high storage costs have a larger increase in local employment, which comes from increases in the number of establishments operating locally, rather than from increases in the size of the establishment. These findings are compatible with a simple theoretical framework where merchants are aware of consumers' desire to economize on travel time, and hence substitute land with labor at a faster rate for high storage-cost sectors in response to higher demand. They appear robust to measures that capture variation in likely consumers only and variation in consumption patterns partially driven by commuting flows.

Our results are subject to a number of caveats that arise from the limited nature of our data and from the attempt to bring under a unified logic consumption markets that are potentially quite different. On the other hand, our findings describe broad patterns of consumer behavior for a large portion of economic

[51] A natural way fo trip chaining to occur is via "shopping centers". A shopping center is "a group of architecturally unified commercial establishments built on a site that is planned, developed, owned, and managed as an operating unit related in its location, size, and type of shops to the trade area that the unit serves." Shopping centers accounted for around 28% of total consumer expenditure in 2005, the latest year available (see Table 1061, Section 22, Statistical Abstract of the United States, 2012; and Consumer Expenditures in 2005, U.S. Dept. of Labor, Bureau of Labor Statistics, report 998).

activity in modern economies. Taken together, they suggest the importance of storability/durability to consumer behavior and that consumer mobility may play an important role in the formation of local equilibrium outcomes. Hence, it may be fruitful to incorporate these margins in future analyses of the consequence of local shocks and place-based policies. Our results also provide important insights for the evaluation of competitive consequences of horizontal mergers, and for firm entry following liberalization of international trade and investment in services: firms' choices appear to respond to different degrees of localization for their product's demand.

References

[1] Alchian, A., and W. Allen (1964). “University Economics”. Belmont, Calif.: Wadsworth.

[2] Agarwal, S., N. Marwell, and L. McGranahan (2017). “Consumption Responses to Temporary Tax Incentives: Evidence from State Sales Tax Holidays”, American Economic Journal: Economic Policy, 9:4, pp. 1–27.

[3] Alessandria, G., J. Kaboski and V. Midrigan (2010). “Inventories, Lumpy Trade, and Large Devaluations,” American Economic Review, Vol. 100:5, pp. 2304-2339.

[4] Anderson, J. (1979). “A Theoretical Foundation for the Gravity Equation”, American Economic Review, Vol. 69:1, 106:116.

[5] Anderson, J. and E. Van Wincoop (2003). “Gravity with Gravitas: a Solution to the Border Puzzle”, American Economic Review, Vol. 93:1, pp. 170-192.

[6] Baker, S., S. Johnson and L. Kueng (2018). “Shopping for Lower Sales Tax Rates,” NBER Working Paper 23665.

[7] Baumol, W. (1952). “The Transactions Demand for Cash: An Inventory Theoretic Approach,” Quarterly Journal of Economics, Vol. 66:4, pp. 545-556.

[8] Brainard, S. L. (1997). "An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade," American Economic Review, Vol. 87:4, pp. 520-544.

[9] Burchfield, M., H. G. Overman, and M. A. Turner (2006). “Causes of Sprawl: A Portrait from Space,” Quarterly Journal of Economics, Vol. 121:2, pp. 587-633.

[10] Cameron, A. C. and K. P. Trivedi (2005). “Microeconometrics: Methods and Applications”, Cambridge University Press.

[11] Chandra, A., K. Head and M. Tappata (2014). “The Economics of Cross-Border Travel”, The Review of Economics and Statistics Vol. 96:4, pp. 648-661.

[12] Chaney, T. (2018). “The Gravity Equation in International Trade: An Explanation”, Vol. 126:1, pp. 150-177.

[13] Christaller, Walter (1933): Die zentralen Orte in Süddeutschland. Gustav Fischer, Jena.

[14] Coleman, A.M.G. 2009. “Storage, slow transport and the law of one price: theory with evidence from nineteenth century U.S. corn markets,” Review of Economics and Statistics 91(2) 332-350.

[15] Combes, P.F., G. Duranton, L. Gobillon, D. Puga and S. Roux (2012). “The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection,” Econometrica, Vol. 80, No. 6, pp. 2543–2594

[16] Costinot, A. and J. Vogel (2010). "Matching and Inequality in the World Economy", Journal of Political Economy, Vol. 118:4, pp. 747-786.

[17] Couture, V. (2016). "Valuing the Consumption Benefits of Urban Density", Working Paper.

[18] Disdier, A. and K. Head (2008). "The Puzzling Persistence of the Distance Effect on Bilateral Trade", Review of Economics and Statistics, Vol. 90:1, pp. 37-48.

[19] Davis, D., J. Dingel, J. Monras, and E. Morales (2018). "How Segregated is Urban Consumption?", Accepted, Journal of Political Economy.

[20] Davis, P. (2006). "Spatial Competition in Retail Markets: Movie Theaters", RAND Journal of Economics, Vol. 37:4, pp. 964-982.

[21] Dornbush, R., S. Fischer and P. Samuelson (1977). "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods", American Economic Review, Vol. 67:5, pp. 823-839.

[22] Duranton, G. M. A. Turner (2017). "Urban form and driving: Evidence from US cities," forthcoming, Journal of Urban Economics.

[23] Eaton, J., and S. Kortum (2003). "Technology, Geography, and Trade", Econometrica, Vol. 70:5, pp. 1741-1779.

[24] Einav, L., P. Klenow, B. Klopack, J. Levin, L. Levin, and W. Best (2017). "Assessing The Gains from E-Commerce", Working Paper.

[25] Ellison, G. and S. F. Ellison (2009). "Tax Sensitivity and Home State Preferences in Internet Purchasing", American Economic Journal: Economic Policy, Vol. 1:2, p. 53-71.

[26] Fafchamps, M. and F. Shilpi, (2003). "The Spatial Division of Labor in Nepal", Journal of Development Studies, Vol. 39:6, pp 23-66.

[27] Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011). "Income Distribution, Product Quality, and International Trade", Vol. 119:4, pp. 721-765.

[28] Fieler, A. C. (2011). "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation", Econometrica, Vol. 79:4, pp. 1069-1101.

[29] Gervais, A. and J.B. Jensen (2018). "The Tradability of Services: Geographic Concentration and Trade Costs," Journal of International Economics, forthcoming.

[30] Glaeser, E., J. Kolko, and A. Saiz (2001). "Consumer City", Journal of Economic Geography, Vol. 1:1, pp. 27-50.

[31] Handbury, J., I. Rahkovsky, and M. Schnell (2018). "Is the Focus on Food Deserts Fruitless? Retail Access and Food Purchases Across the Socioeconomic Spectrum", Working paper.

[32] Hendel, I. and A. Nevo, (2006a). "Sales and consumer inventory," RAND Journal of Economics, Vol. 37, No. 3, pp. 543-561.

[33] Hendel, I. and A. Nevo (2006b). "Measuring the Implications of Sales and Consumer Inventory Behavior," Econometrica, Vol. 74, No. 6, pp. 1637–1673.

[34] Hillberry, R. and D. Hummels (2007). "Trade Responses to Geographic Frictions: a Decomposition Using Micro-Data", European Economic Review Vol. 52:3, pp. 527:550.

[35] Horstmann, I. and J. R. Markusen (1992). "Endogenous Market Structures in International Trade", Journal of International Economics Vol. 32:1-2, pp. 109-129.

[36] Hummels, D. and P. Klenow (2005). "The Variety and Quality of a Nation's Exports", American Economic Review, Vol. 95:3, pp. 704-723.

[37] Hummels, D. and A. Skiba (2004). "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture", Journal of Political Economy, Vol. 112:6, pp. 1384-1402.

[38] Hortaçsu, A., F. A. Martínez-Jerez, and J. Douglas (2009). "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre", American Economic Journal: Microeconomics, Vol. 1:1, p. 53-74.

[39] Hortaçsu, A. and C. Syverson (2015). "The Ongoing Evolution of US Retail: A Format Tug-of-War", Journal of Economic Perspectives, Vol. 29:4, pp. 89-112.

[40] Houde, J.F. (2012). "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline", American Economic Review, Vol. 102:5, pp. 2147-2182.

[41] Jarmin, R., S. Klimek, and J. Miranda (2005). "The Role of Retail Chains: National, Regional and Industry Results", in "Producer Dynamics: New Evidence from Micro Data", University of Chicago Press, p. 237-262.

[42] Karádi, P. and M. Koren (2017). "Cattle, Steaks and Restaurants: Development Accounting when Space Matters," Working paper.

[43] Kline, P. and E. Moretti (2014). "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," Annual Review of Economics, Vol. 6, pp. 629-662.

[44] Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston (2012). "An Overview of the Global Historical Climatology Network-Daily Database", Journal of Atmospheric and Oceanic Technology, Vol. 29, pp. 897-910.

[45] Mian, A. and A. Sufi (2014). "What Explains the 2007-2009 Drop in Employment?", Econometrica, Vol. 82:6, pp. 2197–2223.

[46] Monte, F. (2011). "Skill Bias, Trade, and Wage Dispersion", Journal of International Economics, Vol. 83:2, pp. 202-218.

[47] Monte, F., S. Redding and E. Rossi-Hansberg (2018). "Commuting, Migration, and Local Employment Elasticities", American Economic Review, Vol 108:12, pp. 3855-90.

[48] Murray, K., F. Di Muro, A. Finn and P. Popkowski Leszczyc, (2010). "The Effect of Weather on Consumer Spending," Journal of Retailing and Consumer Services, Vol. 17, pp. 512-520.

[49] Nakamura, E. (2008). "Pass-Through in Retail and Wholesale", American Economic Review: Papers and Proceedings, Vol. 98:2, 430-437.

[50] Oi, W. (1992). "Productivity in the Distributive Trades: the Shopper and the Economies of Massed Reserves", in "Output Measurement in the Services Sectors", University of Chicago Press, p. 161-193.

[51] Raval, D. and T. Rosenbaum (2018). "Why is Distance Important for Hospital Choice? Separating Home Bias from Transport Costs", Working Paper.

[52] Reilly, W. (1931). "The Law of Retail Gravitation", New York: Knickerbocker Press.

[53] Sattinger, M. (1993). "Assignment Models of the Distribution of Earnings," Journal of Economic Literature Vol. 31:2, pp. 831-880.

[54] Schiff, N. (2015). "Cities and Product Variety: Evidence from Restaurants", Journal of Economic Geography, Vol 15, pp. 1085-1123.

[55] Shoag, D. and S. Veuger (2017). "Shops and the City: Evidence on Local Externalities and Local Government Policy from Big-Box Bankruptcies," The Review of Economics and Statistics. Forthcoming.

[56] Von Thunen, J. H. (1966). "Isolated State: an English Edition of Der Isolierte Staat" Oxford: Pergamon Press.

[57] Wang, Z. and A. Wolman (2016). "Payment Choice and Currency Use: Insights from Two Billion Retail Transactions," Journal of Monetary Economics, Vol. 84, pp. 94-115.

[58] Williams, J.C., and B. D. Wright, 1991, Storage and Commodity Markets, Cambridge University Press, Cambridge, U.K.

Web Appendix for "The Geography of Consumption" (Not for Publication)

Sumit Agarwal
National University of Singapore

J. Bradford Jensen
Georgetown University

Ferdinando Monte
Georgetown University

A Data Processing

A.1 Merchant codes

The transaction data classifies merchants using the MCC classification. Classifications of merchants come at a "broad" and "narrow" level. We exclude narrow merchant categories that either refer to a transaction which can be executed without involving physical movement of a provider or a customer, or those that are of commercial, rather than private, nature. These categories broadly include items like airlines, cruise lines, direct marketers, online marketers, insurance, financial institutions, business services, political organizations, and other codes reserved for cash advances and balance transfers. The result is a classification of 27 broad categories.

A.2 Transaction data

The raw transaction data comes from U.S. credit card statements issued between March and October, 2003. Some earlier transactions still appear in the file as the date in which they are recorded, which may not necessarily be the date of the transaction. There are originally 3,530,027 records in the data for 134,008 unique accounts. Each record comes from a line in an individual credit card statement. A record contains the account number, transaction date and post date, amount and type of the transaction, the original merchant category code (MCC), and string information on the merchant name and location. After merging this data with the merchant codes above, 1,247,438 transactions are dropped. Of these dropped observations, around 1.1 million records are related to 1) cash advances, interest, late fees, account adjustments, balance transfers, card payments and similar activities not generated by actual purchases; 2) direct marketers and telemarketers, 3) unknown merchants. We also exclude transactions relating to educational services (where the account-holder likely pays for somebody else), and transportation services and vehicle rentals, where location of transaction and location of service use are different. We further keep only records that are actual purchases ("transaction type" code equal to 253) originating on or after February 1, 2003. This leaves us with 2,156,978 transactions from 80,087 accounts.

A.3 Account data

The account data for the months of March to October 2003 originally comprises 2,272,825 records for 249,032 accounts. Among other things, each line contains the record date (year and month) for the entry, the account number, a person ID, the date of birth and gender of the account holder, an external status code, a reported income, a 5 digit ZIP code and the state of residence. Different lines for the same account may be present in the account data because of various events that affect the account (the end of the billing cycle or updates to the month end balance, for example). 28,928 observations appear to be of inactive cards (no information for state, ZIP code, and date of birth), so we drop them. Towards matching the account information with the transaction data, we start by keeping unique combinations of account number, date of birth, state, ZIP code and record date. We find 4 accounts for which the date of birth of the account holder changes, and we make that information consistent by picking the oldest date of birth. After this adjustment, almost all records are unique within account number-event date. We drop three accounts, where the same set of several ZIP codes are reported for each record date, making it difficult to find a residence location. This processing leaves us with 1,746,667 account number-event date records for 239,369 unique accounts. This step tells us the residence location of an account whenever an account event occurs. Next, we reconstruct where the account holder resided for each of the transactions described above.

A.4 Matching transactions and account data

We match the transaction and account data to assign a location of residence to each purchase. For a given account, we match the month of the transaction in the first file to the event month in the account data, if possible. For those observations where this is not possible, we match the closest account information that precedes the transaction; when this second option is not feasible, we match it with the earliest information following the transaction. The matching process leaves us with 2,138,575 transactions matched from 78,418 unique accounts. Out of the totality of matched transactions, only 151,725 did not find the exact event month in the account information: 142,520 records among these come from transactions in February 2003, which are then matched with information in March.

A.5 Extracting merchant location name

The data provides us with a full merchant name string (including usually merchant name, location/phone number and state) and a merchant name string. Here we explain how we extract the potential city and state names of each transaction.

We first extract the merchant state. The state of the merchant is typically located at the end of the full merchant name. We extract the last two characters of the merchant name string if the last three start with a space. Only 1,588 transactions do not meet this requirement: in most cases, the last two letters still represent a state (or a foreign country), but we won't be able to rule out false positives. We match these states with a list of U.S. states and country abbreviations to verify that we have extracted

U.S. states. We match only 52% of the 1,588 thousand problematic observations, and more than 98% of the other transactions. Keeping only transactions where a U.S. state could be identified leaves us with 2,106,552 observations.

To identify the set of observations we might match with a location name, we start by extracting a potential location name. To do so, we remove from the full merchant name string the merchant name that the data provides (from the left of the string) and the state we have extracted (from the right of the string). This procedure generates 7,777 observations with an empty potential location name.

We then mark transactions of common online providers[52] and find the words "Online", "On Line", ".com", ".net" in 100,265 observations. We mark observations where the final part of the string before the state is a phone number – these are typically online stores – and find 188,316 of them. We are left with 1,901,658 transactions that may contain city names, 90% of those for which a state name could be found, for 73,385 unique accounts. Note that the largest contributor to the drop in observations is transactions with a phone number rather than a location at the end of the merchant name. We will attempt to match this list of location names with a list of U.S. city and place names from the U.S. Census. Before turning to the different steps in that match, we will discuss briefly how we recover the list of cities.

A.6 List of cities and places in the United States

We construct a list of city names and states from the year 2000 U.S. Census Gazetteer List of Places and the year 2000 U.S. Census list of County Subdivisions. The List of Places contains incorporated places and unincorporated Census Designated Places (CDP); it excludes towns in the New England states, New York, and Wisconsin, and boroughs in New York (treated as Minor Civil Divisions, or MCDs). The list of County Subdivisions contains, among other things, MCDs (called for example townships, parishes, districts), and Census County Divisions. Both lists contain, among other things, population in 2000 and latitude and longitude of the location.

While FIPS codes are unique, our match to merchants will be on a location name. Hence it may happen that within the list, we have more than one record with the same name (for example, we may have “Mountain View city” and “Mountain View, CDP”). In those cases, we attribute to a name the coordinates with the highest population in 2000.[53]

A.7 Finding location names in the transactions data and computing distances

We attempt to find the name of a city in four passes. First, we match the location name and state identified above with the list of U.S. Places. We immediately find a match for 1,454,166 out of the 1,901,658 we intend to match, 76% of our observations. Out of the 447,492 transaction with no match, 122,737 have names and states that match the MCD list. We assign "match quality" equal to 10 to those transactions

[52] We identify Paypal, QVC, AOL, Shutterfly, MUI Movies Unlimited, Amazon, Microsoft, Expedia, Untd.com, Ebay, and Netflix.

[53] An alternative could have been to compute the average longitude and latitude of all the occurrences, weighted by population. However, we would still need a unique FIPS code identifier, since accounts will be associated to place codes, not names. This difference makes the approach infeasible.

matched at this first pass. We have 324,755 transactions with no location information (about 17% of the transactions) that we cannot match exactly.

In several instances, the name of a city in the transaction data is truncated from the original. The second pass of the match involves matching truncated versions of city names from the U.S. Census to location names in the transaction data. We assign "match quality" equal to 9 to those cases where the name of a location in the transaction data, of length n, matches the first n characters of a city name. We further assign "match quality" equal to 8 where, for a location name of length n, there is a match in the first $n-1$ characters. Obviously, it can happen that one city in the transaction data can be matched to more than one city in the Census list. We only keep cases where the match is either unique or there are two matches. We solve the two-matches case as follows: if the match is to a Census place and to a minor civil division, we keep the coordinates of the Census place; otherwise, we take the place with the highest population and downgrade the "match quality" by 1. With the second pass, we are able to recover 114,056 observations.

In other instances, some locations may not be matched because of extra spaces or special characters (e.g., "St. Louis" vs. "St Louis"). In the third pass, we "standardize" the name of the remaining unmatched locations by removing all spaces, commas, full stops, and dashes both in the transaction and in the Census files. We assign "match quality" equal to 9 to these observations. With this process, we recover additional 20,796 observations, bringing the number of matched transactions to 1,711,755.

Finally, we identify the remaining unmatched locations with at least one thousand transactions and fix those matches by hand. There are 44 of these instances. We recover 31,664 observations more (also assigned "match quality" equal to 10), bringing the total to 1,743,419 matched transactions, or 91.7% of the transactions we intended to match. For these matched transactions we can attribute a latitude and longitude of the merchant.

The account data provides ZIP code information for each account. We match these ZIP codes against Census Places and (if we don't find a match) MCD lists using concordances for the year 2000 provided by the census. For the few cases in which we cannot find a correspondence, we use analogous ZIP-places and ZIP-MCD concordances for the year 2010. In some cases, a ZIP code may span two or more geographical units: we keep in that case the unit that accounts for the highest fraction of population of the ZIP code. We then have analogous geographies for account and merchant sides, and can compute the bilateral distance between the centroid of the account and shopping locations for each transaction.

The process of matching ZIP codes to geographical areas leads to a small loss in observations. Our working sample has 1,722,873 transactions (90.6% of the transactions we intended to match) and 71,377 accounts. In our classification, 92.2% of observations have match quality equal to 10, and 7.2% have match quality 9, leaving less than 1% of observations with quality 8 (0.61%) and 7 (0.01%).

B Theoretical derivations

B.1 Equilibrium Price

To pin down the constants of integration, we use implications of our conjectured land allocation. In particular, since $j(t)$ is decreasing, the person with the lowest travel cost, $t = 1$, will travel the maximum distance, $j_{\max} \equiv j(1)$. At that distance, the price of the product will have to be zero (otherwise, some firms would have an incentive to enter slightly farther).

Using (17), it follows that

$$-c_1 \exp\{\alpha j_{\max}\} = c_2 \exp\{-\alpha j_{\max}\} \tag{B.1}$$

Using the same information in the implicit function for the distance traveled, when $t = 1 \Longrightarrow j(1) = j_{\max}$, and hence,

$$\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} = c_2 \exp\{-\alpha j_{\max}\} - c_1 \exp\{\alpha j_{\max}\} \tag{B.2}$$

We can substitute (B.1) in the last equation to obtain,

$$\begin{aligned} \left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} &= c_2 \exp\{-\alpha j_{\max}\} + c_1 \exp\{-\alpha j_{\max}\} \\ c_2 &= \frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{\alpha j_{\max}\} \end{aligned} \tag{B.3}$$

and hence

$$c_1 = -\frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{-\alpha j_{\max}\}$$

We can then rewrite the price function as

$$\begin{aligned} p(j) &= -\frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{-\alpha j_{\max}\} \exp\{\alpha j\} + \frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{\alpha j_{\max}\} \exp\{-\alpha j\} = \\ &= \frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \left[\exp\{\alpha (j_{\max} - j)\} - \exp\{-\alpha (j_{\max} - j)\}\right] \end{aligned} \tag{B.4}$$

and the implicit equation for distance,

$$t = \frac{1}{2} \exp\{\alpha (j_{\max} - j)\} + \exp\{-\alpha (j_{\max} - j)\} \tag{B.5}$$

The individual with the highest t travels as close as possible, i.e., $t = 2 \Longrightarrow j(2) = 0$. Imposing this in the equation for distance traveled, the value of $j_{\max}$ is then implicitly defined by

$$4 - \exp\{\alpha j_{\max}\} = \exp\{-\alpha j_{\max}\}$$

For $j_{\max} \geq 0$, the LHS starts at 3, is decreasing and concave, and crosses zero once; the RHS starts at 1, is decreasing and convex, and never crosses zero. Hence, there is a unique solution $j_{\max}(\alpha)$. Totally differentiating this equation with respect to $j_{\max}$ and α,

$$\frac{dj_{\max}}{d\alpha} = -\frac{j_{\max}}{\alpha} < 0$$

This implies that less land is used if α is higher, and also that the elasticity of $j_{\max}$ to α is -1, i.e., the product $\alpha \cdot j_{\max}(\alpha)$ is a constant independent of α.

B.2 Aggregate Implications

Lemma 2 . *The average slope of the expenditure function $X(j)$ between $j \in [0, j_{\max}]$ (unweighted, weighted by the number of agents $n(j)$, and weighted by total expenditure $X(j)$) grows more negative when g is higher.*

Proof. Recall from (19) that

$$p(j) = \frac{1}{2^{1/2}\alpha_0} \cdot \bar{q}^{1/4} g^{1/4} N^{1/2} \cdot [\exp\{\alpha(j_{\max} - j)\} - \exp\{-\alpha(j_{\max} - j)\}], with \bar{p} \equiv \frac{1}{2}\left(\frac{2}{\bar{q}}\right)^{1/2}$$

Differentiating with respect to j, we have

$$p'(j) = -\left(\frac{g}{2\bar{q}}\right)^{1/2} \cdot [\exp\{\alpha(j_{\max} - j)\} + \exp\{-\alpha(j_{\max} - j)\}]$$

The expenditure function $X(j)$ has slope

$$X'(j) = 2\bar{x} \cdot p(j)\, p'(j)$$

The average unweighted slope over $j \in [0, j_{\max}]$ is

$$\begin{aligned} \frac{1}{j_{\max}} \int_0^{j_{\max}} X'(j)\, dj &= \frac{2\bar{x}}{j_{\max}} \int_0^{j_{\max}} p(j)\, p'(j)\, dj = -\left(\frac{g}{2\bar{q}}\right)^{1/2} \cdot \left(\frac{1}{2^{1/2}\alpha_0} \cdot \bar{q}^{1/4} g^{1/4} N^{1/2}\right) \cdot 2\bar{x} \frac{2\sinh[\alpha \cdot j_{\max}]^2}{\alpha \cdot j_{\max}} = \\ &= -\frac{2\sinh[\alpha \cdot j_{\max}]^2}{\alpha \cdot j_{\max}} \frac{\bar{x}}{\alpha_0} \cdot \bar{q}^{-1/4} g^{3/4} N^{1/2} \end{aligned}$$

which is more negative when g is higher. The agents density–weighted average slope is

$$\begin{aligned}
&\frac{1}{j_{\max}}\int_0^{j_{\max}} n\left(j\right)X'\left(j\right)dj=\frac{2\bar{x}}{j_{\max}}\int_0^{j_{\max}} n\left(j\right)p\left(j\right)p'\left(j\right)dj=\frac{2\bar{x}}{j_{\max}}\int_0^{j_{\max}} n\left(j\right)p\left(j\right)p'\left(j\right)dj\\
=&\ \frac{2\bar{x}}{j_{\max}}\left(\frac{\bar{q}}{2g}\right)^{1/2}\alpha^2\int_0^{j_{\max}} p\left(j\right)^2p'\left(j\right)dj=\frac{2\bar{x}}{j_{\max}}\left(\frac{\bar{q}}{2g}\right)^{1/2}\alpha^2\int_0^{j_{\max}} p\left(j\right)^2p'\left(j\right)dj=\\
=&\ -\left(\frac{g}{2\bar{q}}\right)^{1/2}\frac{2\bar{x}}{j_{\max}}\left(\frac{\bar{q}}{2g}\right)^{1/2}\alpha^2\left(\frac{1}{2^{1/2}\alpha_0}\cdot\bar{q}^{1/4}g^{1/4}N^{1/2}\right)^2\frac{8\sinh\left[\alpha\cdot j_{\max}\right]^3}{3\alpha}=\\
=&\ -\frac{8\sinh\left[\alpha\cdot j_{\max}\right]^3}{3\alpha\cdot j_{\max}}\bar{x}\left(\frac{\alpha_0 g^{1/4}}{N^{1/2}\bar{q}^{3/4}}\right)^2\left(\frac{1}{2^{1/2}\alpha_0}\cdot\bar{q}^{1/4}g^{1/4}N^{1/2}\right)^2=\\
=&\ -\frac{4\sinh\left[\alpha\cdot j_{\max}\right]^3}{3\alpha\cdot j_{\max}}\bar{x}\frac{g}{\bar{q}}
\end{aligned}$$

which is more negative when g is higher. The expenditure–weighted average slope is

$$\begin{aligned}
&\frac{1}{j_{\max}}\int_0^{j_{\max}} X\left(j\right)X'\left(j\right)dj=\frac{2\bar{x}^2}{j_{\max}}\int_0^{j_{\max}} p\left(j\right)^2p'\left(j\right)dj=\\
=&\ -\frac{2\bar{x}^2}{j_{\max}}\left(\frac{g}{2\bar{q}}\right)^{1/2}\left(\frac{1}{2^{1/2}\alpha_0}\cdot\bar{q}^{1/4}g^{1/4}N^{1/2}\right)^2\frac{8\sinh\left[\alpha\cdot j_{\max}\right]^3}{3\alpha}=\\
=&\ -\frac{\alpha_0^2}{2^{1/2}}\frac{4\sinh\left[\alpha\cdot j_{\max}\right]^3}{3\alpha\cdot j_{\max}}\cdot Ng
\end{aligned}$$

which is again more negative if g is higher. ■

Lemma 3 . *The frequency of trips increases with g for each individual.*

Proof. Since the frequency of trips is $\bar{q}/z$, we consider the behavior of the batch size. We show that agents buy smaller batches as g grows, implying they travel more frequently. From (11) and using the functional form assumptions, the batch size is

$$\tilde{z}\left(t;j\right)\equiv z\left(j\left(t;g\right),t\right)=\left(\frac{2\bar{q}\cdot tj\left(t;g\right)}{g}\right)^{1/2}$$

where we have evaluated the batch at the optimal distance for agent t, $j\left(t\right)$, and we have made the dependence of the travel function on the parameter g explicit. As g grows, the optimal batch for given distance $j\left(t;g\right)$ shrinks via the denominator. Also, the optimal distance traveled $j\left(t;g\right)$ falls with g for every agent. To see this, recall that $j\left(t\right)$ is implicitly defined by (B.5). Totally differentiating with respect to j and t,

$$dt=-\left(\frac{\bar{q}}{2g}\right)^{1/2}p''\left(j\right)dj\Longrightarrow\frac{dj}{dt}=-\left(\frac{2g}{\bar{q}}\right)^{1/2}\frac{1}{p''\left(j\right)}\Longrightarrow j'\left(t\right)=-\left(\frac{2}{\bar{q}}\right)^{1/2}\frac{N}{\alpha_0^2p\left(j\left(t\right)\right)}$$

where we have used $p''\left(j\right)=\alpha^2p\left(j\right)$ from (16) and the definition of α. Differentiating with respect to t, one can verify that $j(t)$ is always convex:

$$j''(t) = +\left(\frac{2}{\bar{q}}\right)^{1/2} \frac{N}{\alpha_0^2 p(j)^2} p'(j)\, j'(t) > 0$$

since $p' < 0$ and $j' < 0$. Consider the function $j(t;g)$ in the space (t,j), for two values $g_1 < g_2$. Since $j_{\max}$ is decreasing in g, $j(1;g_1) > j(1,g_2)$, that is, $j(t;g)$ starts at a lower value when g is higher. Since both curves are always decreasing convex, $j(1;g_1) > j(1,g_2)$ implies that they will cross at most once in $t \in [1,2]$. However, for both values of g, $j(2;g) = 0$; hence, they cannot cross before, and $j(t;g_1) > j(t,g_2)$ $\forall t \in [1,2)$. For any agent t, the distance traveled decreases with g and so the batch size falls. This implies that the frequency of trips increases for every agent. ■

Lemma 4 . *The average distance at which output is produced is*

$$\frac{\int_0^{j_{\max}} jQ(j)\,dj}{N\bar{q}} = \bar{d} \cdot \frac{N^{1/2}}{g^{1/4}}$$

Proof. Using the expression for output (8),

$$\begin{aligned}
\frac{\int_0^{j_{\max}} jQ(j)\,dj}{N\bar{q}} &= \frac{A^2}{2}\left(\frac{\bar{D}}{w}\right)\frac{1}{N\bar{q}}\int_0^{j_{\max}} jp(j)\,dj = \\
= \frac{A^2}{2}\left(\frac{\bar{D}}{w}\right)\frac{1}{N\bar{q}}\frac{1}{2^{1/2}\alpha_0} &\cdot \bar{q}^{1/4} g^{1/4} N^{1/2} \cdot \frac{2\left[\sinh(\alpha j_{\max}) - \alpha j_{\max}\right]}{\alpha^2} = \\
= \frac{A^2}{2}\left(\frac{\bar{D}}{w}\right)\frac{1}{N\bar{q}}\frac{1}{2^{1/2}\alpha_0} &\cdot \bar{q}^{1/4} g^{1/4} N^{1/2}\left(N^{1/2}\bar{q}^{3/4}\right)^2 \cdot \frac{2\left[\sinh(\alpha j_{\max}) - \alpha j_{\max}\right]}{\left(\alpha_0 g^{1/4}\right)^2} = \\
= \frac{\left[\sinh(\alpha j_{\max}) - \alpha j_{\max}\right]}{\alpha_0} &\cdot \bar{q}^{3/4} g^{-1/4} N^{1/2} = \bar{d} \cdot \frac{\bar{q}^{3/4} N^{1/2}}{g^{1/4}}
\end{aligned}$$

■

Lemma 5 . *The equilibrium total employment is*

$$L_{eq} = \bar{L} \cdot g^{1/4} N^{3/2}$$

Proof. Using the expression for labor demand (7) and $\beta = 1/2$

$$
\begin{aligned}
& L_{eq} \equiv \int_0^{j_{\max}} L(j)\, dj = \bar{D} \left(\frac{1}{2} \frac{A}{w} \right)^2 \int_0^{j_{\max}} p(j)^2\, dj = \\
= \quad & \bar{D} \left(\frac{1}{2} \frac{A}{w} \right)^2 \left(\frac{1}{2^{1/2} \alpha_0} \cdot \bar{q}^{1/4} g^{1/4} N^{1/2} \right)^2 \left[\frac{\sinh(2\alpha j_{\max})}{\alpha} - 2 j_{\max} \right] = \\
= \quad & \bar{D} \left(\frac{1}{2} \frac{A}{w} \right)^2 \left[\sinh(2\alpha j_{\max}) - 2 j_{\max} \alpha \right] \left(\frac{1}{2^{1/2} \alpha_0} \right)^2 \frac{q^{5/4} g^{1/4} N^{3/2}}{2^{-1/4} A \left(\bar{D}/w \right)^{1/2}} = \\
= \quad & \frac{\sinh(2\alpha j_{\max}) - 2 j_{\max} \alpha}{2^{3/2} A \bar{D}^{1/2} w^{1/2}} \cdot q^{5/4} g^{1/4} N^{3/2} = \\
= \quad & \frac{\bar{L}}{A w^{1/2}} g^{1/4} N^{3/2}
\end{aligned}
$$

with $\bar{L} \equiv \frac{\sinh(2\alpha j_{\max}) - 2 j_{\max} \alpha}{2^{3/2} A \bar{D}^{1/2} w^{1/2}} q^{5/4}$. Note that $\bar{L}$ does not vary with g or N since $\alpha j_{\max}$ is constant with α.
■

C Additional Empirical Results

C.1 Summary Statistics by state

Table C.1 shows summary statistics on our main dataset by state of transaction.

C.2 Frequent users

Here we focus on consumers with at least 120 transactions in the sample (that is, around 2 transactions per day from March to October). We term this "frequent users" (FUs) sample, and use it to show that the limited mobility of consumers described above does not depend on including low frequency usage. Our FUs sample contains 1,955 accounts, conducting around 377 thousand transactions over the sample period. They reside in 1,399 locations and shop in 6,149 of them; there are a total of 21,650 origin-destination combinations over which we observe transactions.

Table C.2 shows summary statistics for this sample. Consumers in the median residence visit only 13 distinct sales locations overall during the sample period (15.5 sales location on average). Both values are higher than in the complete data; however, these consumers also live in places with richer options: the median residence records 241 sales locations within 120 km (compared to 192 for the whole data). Hence, the median residence sees consumers shop in 5% of the available locations (the mean is 6%), very comparable to the values in the general data (4% and 7% respectively).

Table C.3 replicates Table 3 in the sample of frequent users. The role of distance is twice as high as the role of locations available. The distance elasticity is closer to conventional levels also found in the trade literature.

Table C.1: **Summary of transaction amounts (in USD), by U.S. State of purchase**

State	Median	Mean	St. Dev.	Sum	N
AK	32	69	132	122,111	1,774
AL	28	63	171	1,057,448	16,905
AR	29	62	154	536,710	8,654
AZ	28	69	230	1,768,032	25,681
CA	30	72	207	10,504,912	146,418
CO	26	60	179	1,655,955	27,636
CT	31	68	178	4,047,578	59,444
DC	26	64	163	249,546	3,917
DE	30	72	216	482,253	6,680
FL	30	70	212	7,143,974	102,526
GA	27	63	181	2,621,643	41,767
HI	33	78	205	405,416	5,196
IA	28	60	167	795,665	13,366
ID	29	64	158	298,824	4,671
IL	30	68	181	4,647,933	68,574
IN	29	63	161	2,168,487	34,338
KS	29	62	183	966,213	15,656
KY	29	63	192	1,088,033	17,216
LA	29	61	140	1,151,874	18,865
MA	31	67	166	10,239,352	152,870
MD	28	67	185	2,404,395	35,802
ME	32	70	168	1,161,195	16,553
MI	30	66	166	3,146,431	48,022
MN	29	67	172	1,720,008	25,707
MO	29	65	184	1,859,377	28,612
MS	30	65	186	502,186	7,688
MT	33	65	130	283,635	4,358
NC	28	65	180	2,414,488	37,408
ND	29	63	142	209,849	3,337
NE	30	67	191	519,324	7,707
NH	32	80	274	1,819,532	22,853
NJ	31	71	202	7,149,537	100,840
NM	28	63	193	478,979	7,576
NV	40	90	229	1,169,957	13,033
NY	33	75	194	11,053,563	147,574
OH	29	65	168	3,707,650	57,383
OK	29	65	171	841,117	12,910
OR	28	63	186	1,237,409	19,743
PA	30	67	174	4,719,180	70,287
RI	31	68	167	1,105,141	16,292
SC	28	67	216	1,200,522	17,927
SD	34	73	216	241,186	3,323
TN	29	65	164	1,703,392	26,318
TX	26	61	182	5,919,181	97,279
UT	26	65	238	581,595	8,983
VA	28	65	192	2,881,000	44,257
VT	30	70	166	312,755	4,464
WA	26	64	190	1,481,652	23,134
WI	30	67	209	2,224,750	33,065
WV	31	67	172	384,579	5,741
WY	30	65	161	165,159	2,543
Total	30	68	188	116,550,684	1,722,873

Table C.2: **Summary statistics across residence locations (Frequent Users)**

variable	**min**	**p10**	**p25**	**p50**	**p75**	**p90**	**max**	**mean**	**N**
Sales locations visited	1	6	9	13	20	27	129	15.47	1,399
Sales locations available	8	90	151	241	526	848	1,110	357.47	1,399
Mean distance to sales locations	21.1	59.1	64.9	71.1	76.8	81	95	70.43	1,399
Share available locations visited	0	0.02	0.03	0.05	0.08	0.13	0.46	0.06	1,399

Table C.3: Locations available and locations visited

Dependent variable:	Log of number of sales locations visited		
	(1)	(2)	(3)
Sales locations within 120km, log	0.443***		0.464***
	(0.017)		(0.017)
Average distance to sales locations within 120km, log		-0.619***	-1.014***
		(0.164)	(0.118)
Constant	0.081	5.169***	4.272***
	(0.095)	(0.699)	(0.502)
R^2	0.33	0.02	0.37
N	1,399	1,399	1,399

Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C.3 Percentiles of distances traveled

These Tables show summary statistics on the percentiles of distances traveled by consumers by sector. Table C.4 refers to percentiles in the unweighted distribution. Table C.5 shows the same percentiles weighting each transaction with the correspondent purchase value.

Table C.4: **Distribution of transaction distances (in km), by sector**

	p10	**p25**	**p50**	**p75**	**p90**	**p99**	**max**	**mean**
Agricultural Services	0	0	5.6	14.6	28.7	1,514.1	6,372.1	50.1
Amusement, Rec. Serv.	0	7.9	33.3	327.1	1,600.3	4,130	8,237.4	454.1
Apparel	0	4.7	15.6	52.1	364.7	3,825.3	8,253.1	201.1
Auto Repair/Service/Parking	0	0	7.9	24	78	2,315.3	7,937.3	94.7
Auto and Truck Sales/Service/Parts	0	0	8.3	21	58.6	2,119	7,775.3	88.8
Building Mat./Hardware/Garden Supp.	0	0	7.6	18.2	40.6	1,491.7	7,868.1	49.6
Communications	0	6.5	24.3	684.8	2,018	3,944.5	8,134.9	551.1
Durable Goods	0	5.8	22	162.2	1,652.5	3,946.5	7,115	420.9
Eating and Drinking Places	0	0	12.6	50.4	496.4	3,739.5	8,254.8	217.1
Food Stores	0	0	4.2	15.3	53.8	2,409.1	8,218	93.8
Furniture, Home Furnishings, Equip.	0	1.9	11.2	26.3	132.8	3,277.2	8,243.6	135.9
Gasoline Services	0	0	8.9	34.6	275	2,274.3	8,233.3	126.8
General Merchandise Stores	0	0	8.7	20.8	61.3	2,001.5	8,223.9	87.3
Health Services	0	0	8.6	20.3	46.3	2,231.3	7,969.9	83.9
Hospitality	51.3	162.8	366.8	1,011.1	2,257.8	4,158.5	8,253.1	801.8
Misc. Retail	0	0	8.6	29.6	353.9	3,729.5	8,223.9	192.6
Misc. Services	0	2.2	15.8	67.8	1,131.8	3,905.3	7,765.3	302.3
Motion Pictures	0	0	5.7	16.6	63.6	3,756.9	7,884.2	125.3
NonDurable Goods	0	0	8.2	22	143.7	3,421.9	7,768.4	145
Other Vehicles Sales/Service/Parts	0	6.6	20.1	55	505.9	3,017.7	7,879.4	190.8
Personal Services	0	0	6.7	20	135.1	3,332.5	8,251.4	132.3
Total	0	0	9	29.4	276.8	3,249.4	8,254.8	157.4

Table C.5: **Value-Weighted Distribution of transaction distances (in km), by sector**

	p10	**p25**	**p50**	**p75**	**p90**	**p99**	**max**	**mean**
Agricultural Services	0	0	6.7	16.8	36.3	1,348.7	6,372.1	52
Amusement, Rec. Serv.	0	8.2	37.3	419.5	1,752.7	4,290.5	8,237.4	530.3
Apparel	0	5.3	16.8	56.1	438.5	3,864.5	8,253.1	222.1
Auto Repair/Service/Parking	0	0	7.5	20.3	65.2	2,080.8	7,937.3	86.9
Auto and Truck Sales/Service/Parts	0	0	11.7	27.8	113.3	2,246.5	7,775.3	105.8
Building Mat./Hardware/Garden Supp.	0	0	9.9	23.4	54.2	1,572.5	7,868.1	56.2
Communications	0	4.5	14.7	113.6	1,522	3,818.2	8,134.9	367.8
Durable Goods	0	10.5	30.6	198.1	1,864.6	4,017.6	7,115	454.7
Eating and Drinking Places	0	1	15.4	79.3	711.2	3,940.4	8,254.8	264.2
Food Stores	0	0	5.2	16.9	55	2,374.7	8,218	91.8
Furniture, Home Furnishings, Equip.	0	4.6	13.1	30.6	129.2	2,966.2	8,243.6	129.6
Gasoline Services	0	0	9.7	39.6	320.1	2,247.8	8,233.3	133.9
General Merchandise Stores	0	0	9.9	23.1	77.2	2,547.3	8,223.9	104
Health Services	0	0	9.8	24.9	75.7	2,686.1	7,969.9	112.3
Hospitality	59.6	179.2	434.1	1,320.3	2,664.2	4,331.4	8,253.1	949.2
Misc. Retail	0	0	13	49.7	703.8	3,911.7	8,223.9	254.7
Misc. Services	0	5.3	17.1	54.7	666.7	3,964.9	7,765.3	238
Motion Pictures	0	0	7.3	22.2	222.7	3,960.8	7,884.2	181.4
NonDurable Goods	0	3	11.2	34.1	742.2	3,942.2	7,768.4	249.6
Other Vehicles Sales/Service/Parts	0	9.1	23.1	64.9	936.9	3,139	7,879.4	244.7
Personal Services	0	0	10.9	37.5	529.8	3,856.4	8,251.4	217.2
Total	0	0	12.3	40.5	434.7	3,709.5	8,254.8	205.3

C.4 Gravity over all distances

In Figure C.1, we estimate Equation (2) including origin-destination pairs at progressively longer distances. Specifically, we split all the (h, s) pairs in 20 quantiles of distances, and estimate it using only the first group, then only the first two, and so on, up to the whole set of observations. Figure C.1 shows the coefficient on log distance. As one can see, changes of around +/- 30% in the 120 km cutoff (from 80 km to 160 km) only imply a variation in the gravity coefficient of around 0.1: hence, around our cutoff distance, the overall gravity slope is not particularly sensitive to the specific cutoff value. Different sectors are more or less represented at different distances (see also Tables C.4 and C.5), implying that the coefficient δ varies.

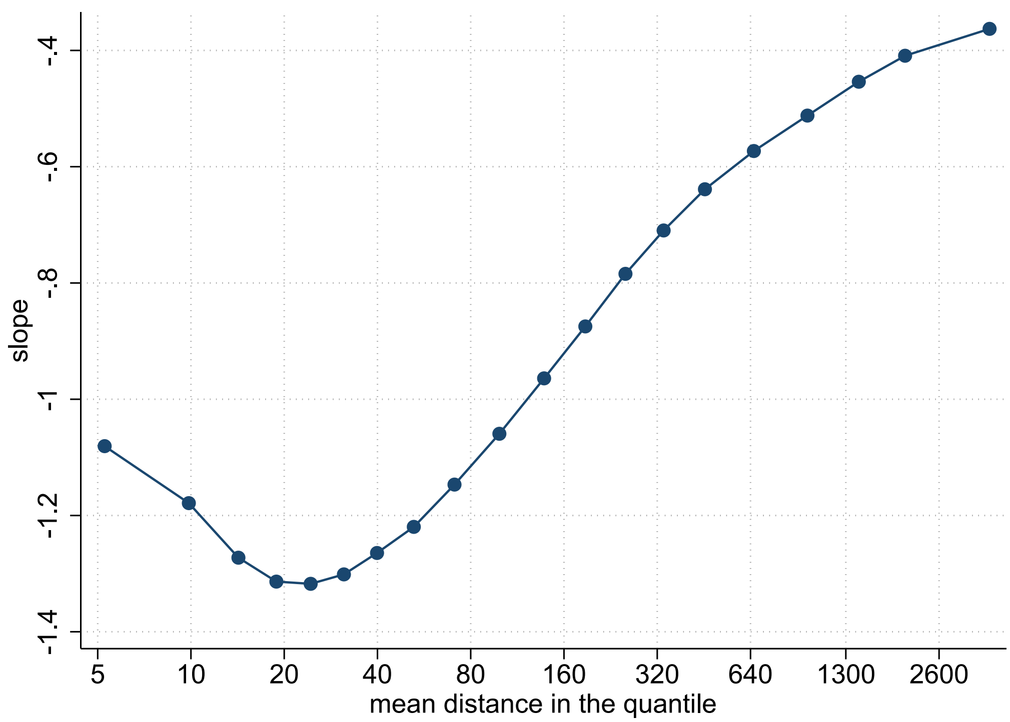

Figure C.1: Gravity in Expenditure

C.5 Margins decomposition

Tables C.6 shows the actual values of the account and expenditure margin with associated p-values for the margins decomposition associated with Equation (1); Table C.7 shows the actual values of the account and expenditure margin with associated p-values associated with Equation (2).

Tables C.8 and C.9 show the composition of frequency and batch size margin into the overall expenditure margin. They also show the share of the frequency margin in the expenditure margin, and the overall role of frequency and account margins in the total decline of expenditure with distance. As in the main text, all p-values are computed using heteroskedasticity–robust standard errors; the number of observations reported excludes singletons, i.e. those observations that would be absorbed by fixed effects and do not contribute to the estimation.

Table C.6: **Expenditure out of home place**

	Overall		Accounts Margin		Expenditure Margin		Share Accounts	Obs.
Category	coeff	pv	coeff	pv	coeff	pv	Margin	
Food Stores	-2.23	0.00	-1.12	0.00	-1.11	0.00	0.50	22,649
Gasoline Services	-2.08	0.00	-0.97	0.00	-1.11	0.00	0.47	39,666
General Merchandise Stores	-1.78	0.00	-1.08	0.00	-0.71	0.00	0.60	26,837
Misc. Retail	-1.70	0.00	-1.07	0.00	-0.63	0.00	0.63	34,052
Eating and Drinking Places	-1.57	0.00	-0.93	0.00	-0.64	0.00	0.59	34,504
Building Mat./Hardware/Garden Supp.	-1.40	0.00	-0.87	0.00	-0.53	0.00	0.62	14,185
Auto Repair/Service/Parking	-1.25	0.00	-0.88	0.00	-0.38	0.00	0.70	4,414
NonDurable Goods	-1.16	0.00	-1.05	0.00	-0.11	0.45	0.91	978
Health Services	-1.12	0.00	-0.77	0.00	-0.35	0.00	0.68	5,134
Apparel	-1.10	0.00	-0.83	0.00	-0.27	0.00	0.75	15,918
Furniture, Home Furnishings, Equip.	-1.07	0.00	-0.85	0.00	-0.23	0.00	0.79	12,286
Auto and Truck Sales/Service/Parts	-1.04	0.00	-0.81	0.00	-0.23	0.00	0.78	7,298
Motion Pictures	-1.04	0.00	-0.85	0.00	-0.18	0.01	0.82	1,922
Amusement, Rec. Serv.	-1.03	0.00	-0.66	0.00	-0.37	0.00	0.64	2,958
Personal Services	-0.96	0.00	-0.89	0.00	-0.07	0.12	0.93	5,203
Misc. Services	-0.92	0.06	-0.63	0.00	-0.29	0.52	0.69	220
Communications	-0.89	0.00	-0.61	0.00	-0.28	0.04	0.69	424
Agricultural Services	-0.88	0.00	-0.66	0.00	-0.21	0.10	0.75	552
Other Vehicles Sales/Service/Parts	-0.68	0.41	-0.71	0.00	0.03	0.97	1.04	257
Hospitality	-0.64	0.01	-0.49	0.00	-0.15	0.40	0.76	1,392
Durable Goods	-0.09	0.90	-0.27	0.04	0.18	0.76	3.15	79

Table C.7: **Gravity in expenditure**

	Overall		Accounts Margin		Expenditure Margin		Share Accounts	Obs.
Category	coeff	pv	coeff	pv	coeff	pv	Margin	
Food Stores	-0.85	0.00	-0.36	0.00	-0.49	0.00	0.42	18,632
Gasoline Services	-0.60	0.00	-0.25	0.00	-0.35	0.00	0.41	34,615
General Merchandise Stores	-0.93	0.00	-0.50	0.00	-0.43	0.00	0.54	23,932
Misc. Retail	-0.65	0.00	-0.40	0.00	-0.25	0.00	0.61	30,042
Eating and Drinking Places	-0.56	0.00	-0.31	0.00	-0.25	0.00	0.55	31,022
Building Mat./Hardware/Garden Supp.	-0.73	0.00	-0.39	0.00	-0.34	0.00	0.53	11,604
Auto Repair/Service/Parking	-0.40	0.00	-0.23	0.00	-0.16	0.00	0.59	3,013
NonDurable Goods	-0.65	0.00	-0.40	0.00	-0.24	0.01	0.62	758
Health Services	-0.33	0.00	-0.25	0.00	-0.09	0.08	0.74	3,910
Apparel	-0.53	0.00	-0.36	0.00	-0.17	0.00	0.67	14,066
Furniture, Home Furnishings, Equip.	-0.57	0.00	-0.40	0.00	-0.17	0.00	0.70	10,734
Auto and Truck Sales/Service/Parts	-0.33	0.00	-0.26	0.00	-0.07	0.08	0.79	5,508
Motion Pictures	-0.34	0.00	-0.28	0.00	-0.07	0.22	0.80	1,248
Amusement, Rec. Serv.	-0.23	0.00	-0.10	0.00	-0.13	0.00	0.44	2,329
Personal Services	-0.31	0.00	-0.27	0.00	-0.04	0.27	0.86	3,760
Misc. Services	0.91	0.02	-0.11	0.06	1.02	0.01	-0.12	116
Communications	-0.41	0.01	-0.26	0.00	-0.15	0.21	0.63	263
Agricultural Services	0.42	0.11	-0.12	0.21	0.54	0.03	-0.28	190
Other Vehicles Sales/Service/Parts	-0.59	0.08	-0.07	0.17	-0.51	0.10	0.13	128
Hospitality	-0.14	0.08	-0.08	0.00	-0.06	0.39	0.55	1,158
Durable Goods	1.11	0.67	0.00		1.11	0.67	0.00	15

Table C.8: **Expenditure out of home place: number of transactions and average expenditure**

	Expenditure margin		Batch size margin		Frequency margin		Share of Frequency	Share of Account+Frequency	Obs.
Category	coeff	pv	coeff	pv	coeff	pv	margin	margins	
Food Stores	-1.11	0.00	-0.18	0.00	-0.93	0.00	0.84	0.92	22,649
Gasoline Services	-1.11	0.00	-0.09	0.00	-1.02	0.00	0.92	0.96	39,666
General Merchandise Stores	-0.71	0.00	-0.06	0.00	-0.65	0.00	0.91	0.97	26,837
Misc. Retail	-0.63	0.00	0.05	0.00	-0.68	0.00	1.08	1.03	34,052
Eating and Drinking Places	-0.64	0.00	0.02	0.05	-0.66	0.00	1.04	1.02	34,504
Building Mat./Hardware/Garden Supp.	-0.53	0.00	-0.02	0.48	-0.51	0.00	0.96	0.99	14,185
Auto Repair/Service/Parking	-0.38	0.00	-0.21	0.00	-0.16	0.00	0.43	0.83	4,414
NonDurable Goods	-0.11	0.45	0.02	0.88	-0.13	0.04	1.17	1.02	978
Health Services	-0.35	0.00	-0.17	0.00	-0.18	0.00	0.52	0.85	5,134
Apparel	-0.27	0.00	-0.01	0.54	-0.26	0.00	0.95	0.99	15,918
Furniture, Home Furnishings, Equip.	-0.23	0.00	-0.01	0.88	-0.22	0.00	0.97	0.99	12,286
Auto and Truck Sales/Service/Parts	-0.23	0.00	-0.01	0.85	-0.22	0.00	0.96	0.99	7,298
Motion Pictures	-0.18	0.01	0.02	0.72	-0.20	0.00	1.10	1.02	1,922
Amusement, Rec. Serv.	-0.37	0.00	-0.19	0.01	-0.18	0.00	0.48	0.81	2,958
Personal Services	-0.07	0.12	0.16	0.00	-0.23	0.00	3.31	1.17	5,203
Misc. Services	-0.29	0.52	-0.20	0.64	-0.09	0.37	0.32	0.79	220
Communications	-0.28	0.04	-0.17	0.25	-0.11	0.09	0.38	0.81	424
Agricultural Services	-0.21	0.10	0.01	0.94	-0.22	0.00	1.04	1.01	552
Other Vehicles Sales/Service/Parts	0.03	0.97	0.27	0.71	-0.24	0.28	-7.96	1.39	257
Hospitality	-0.15	0.40	-0.04	0.81	-0.11	0.11	0.75	0.94	1,392
Durable Goods	0.18	0.76	0.02	0.97	0.17	0.50	0.91	1.19	79

Table C.9: **Gravity in expenditure: number of transactions and average expenditure**

Category	Expenditure margin coeff	Expenditure margin pv	Batch size margin coeff	Batch size margin pv	Frequency margin coeff	Frequency margin pv	Share of Frequency margin	Share of of Account+Frequency margins	Obs.
Food Stores	-0.49	0.00	-0.13	0.00	-0.36	0.00	0.73	0.84	18,632
Gasoline Services	-0.35	0.00	-0.04	0.00	-0.31	0.00	0.89	0.93	34,615
General Merchandise Stores	-0.43	0.00	-0.09	0.00	-0.33	0.00	0.78	0.90	23,932
Misc. Retail	-0.25	0.00	-0.01	0.16	-0.24	0.00	0.95	0.98	30,042
Eating and Drinking Places	-0.25	0.00	-0.02	0.00	-0.23	0.00	0.90	0.96	31,022
Building Mat./Hardware/Garden Supp.	-0.34	0.00	-0.07	0.00	-0.27	0.00	0.80	0.91	11,604
Auto Repair/Service/Parking	-0.16	0.00	-0.09	0.07	-0.07	0.00	0.44	0.77	3,013
NonDurable Goods	-0.24	0.01	-0.09	0.23	-0.15	0.00	0.62	0.86	758
Health Services	-0.09	0.08	0.03	0.56	-0.11	0.00	1.30	1.08	3,910
Apparel	-0.17	0.00	-0.02	0.12	-0.16	0.00	0.90	0.97	14,066
Furniture, Home Furnishings, Equip.	-0.17	0.00	-0.04	0.06	-0.13	0.00	0.77	0.93	10,734
Auto and Truck Sales/Service/Parts	-0.07	0.08	0.02	0.53	-0.09	0.00	1.33	1.07	5,508
Motion Pictures	-0.07	0.22	0.02	0.72	-0.08	0.03	1.23	1.05	1,248
Amusement, Rec. Serv.	-0.13	0.00	-0.04	0.28	-0.08	0.00	0.66	0.81	2,329
Personal Services	-0.04	0.27	0.08	0.02	-0.12	0.00	2.84	1.25	3,760
Misc. Services	1.02	0.01	1.13	0.00	-0.11	0.18	-0.11	-0.24	116
Communications	-0.15	0.21	-0.24	0.05	0.09	0.15	-0.61	0.40	263
Agricultural Services	0.54	0.03	0.68	0.01	-0.15	0.35	-0.27	-0.63	190
Other Vehicles Sales/Service/Parts	-0.51	0.10	-0.51	0.10	-0.00	0.98	0.01	0.13	128
Hospitality	-0.06	0.39	-0.05	0.41	-0.01	0.72	0.18	0.63	1,158
Durable Goods	1.11	0.67	0.96	0.73	0.15	0.79	0.14	0.14	15

C.6 Gravity and the frequency of transactions

These figures show further robustness on the relation between gravity and the frequency of transactions. Figure C.2 shows the correspondent of Figure 3 using all coefficients, not just the ones significantly different from zero; one can clearly note the outlier "Durable Goods" in the top-left part of the graph. Figure C.3 uses the strength of gravity as measured by regression (2) using all estimated slopes. We have also experimented with an alternative measure of frequency that gives more weight to users that spend more overall, with essentially identical results.

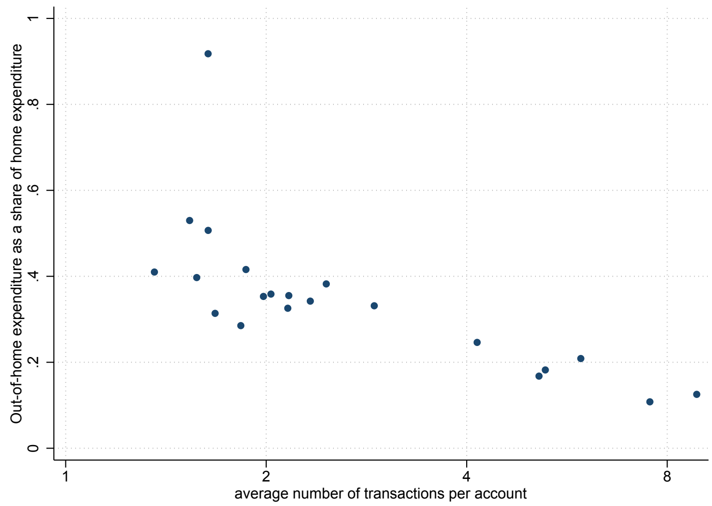

Figure C.2: Drop in expenditure out of home (all coefficients)

C.7 More individual–level analyses

In this subsection we report robustness exercises on our individual level analysis. Table C.10 extends Table 5 in the main text, adding the economy–wide average number of employees per store for different sectors (a proxy for sector–level fixed costs) and its interactions with log age and log income. The interaction between log frequency of transactions with log income and age are little affected both in magnitude and in significance.

A well–known property of the Poisson model is equidispersion: if the mean of the Poisson random variable y_i for individual i is $E[y_i] \equiv \lambda_i = \exp\{\beta' x_i\}$, then $V[y_i] = \exp\{\beta' x_i\}$ as well. This is potentially problematic since count data tend to be overdispersed. Without correction, a strict estimate of the Poisson model will tend to underestimate the standard errors of our coefficients.[54] One possible correction to this problem is to simply estimate the Poisson regression allowing more flexible specifications for the standard errors: a Poisson pseudo–maximum likelihood estimation is still consistent under correct specification of the conditional mean. This is the route we choose in the main text. An alternative is to explicitly model overdispersion. Here, we use a Negative Binomial model where $E[y_i] \equiv \lambda_i$, but $V[y_i] = \lambda_i + \alpha \lambda_i^2$, with α being an additional overdispersion parameter to be estimated. An attractive feature of this model is that the overdispersion is allowed to vary at individual level, since $V[y_i]/E[y_i] = 1 + \alpha \lambda_i$; on the other hand, the Negative Binomial model is less robust to mis-specifications in density,[55] and potentially suffers from the incidental parameter problem, making it difficult to account for individual level heterogeneity[56]. For these reasons, we focus on Poisson in the main text, and report in Table C.11 below the results of the Negative Binomial model.

[54] The discussion here follows Cameron and Trivedi (2005).

[55] See again Cameron and Trivedi (2005), section 20.4.1.

[56] Allison and Waterman (2002) find in simulation studies that the incidental parameter problem may not be very severe.

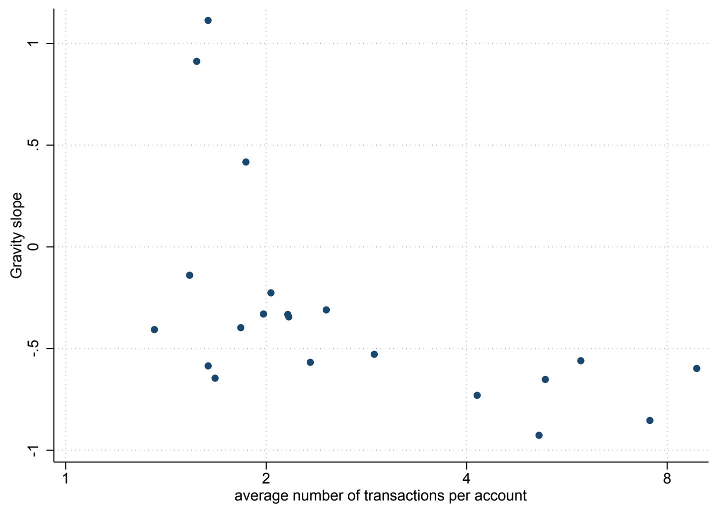

Figure C.3: Gravity and frequency of transactions (all slopes)

Table C.12 extends Table 6 in the main text, introducing the economy–wide average number of employees per store for different sectors, and its interactions with log age and log income. Column (7), additionally, replicates column (6) introducing triple interactions between rain, log frequency, and log income (or log age); and rain, log number of employees per store, and log income (or log age). Table C.13 replicates Table C.12 estimating a Negative Binomial model.

Table C.10: **The role of individual heterogeneity (extended regression)**

	(1)	(2)	(3)	(4)
Dependent Variable:	Number of transactions out of residence			
Number of transactions	0.016***	0.016***	0.022***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
Log age	-0.026	-0.395**	-0.271*	
	(0.059)	(0.159)	(0.152)	
Log income	0.139***	0.589***	0.390***	
	(0.029)	(0.070)	(0.067)	
Log age × log frequency of transactions		-0.027	-0.055	-0.052
		(0.061)	(0.057)	(0.057)
Log income × log frequency of transactions		-0.203***	-0.197***	-0.192***
		(0.029)	(0.025)	(0.025)
Log age × log of employees per store		0.155***	0.117***	0.115***
		(0.040)	(0.030)	(0.030)
Log income × log of employees per store		-0.040**	-0.050***	-0.051***
		(0.019)	(0.013)	(0.013)
Observations	28,959	28,959	28,959	28,959
Sector fixed effects	Yes	Yes	Yes	Yes
ZIP code fixed effects	No	No	Yes	No
Individual fixed effects	No	No	No	Yes
Pseudo R-Square	.61	.61	.74	.74

Standard errors clustered at account level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

C.8 More on density

In this subsection, we examine the role of density and sorting on preferences. As described in the main text, one might imagine that aquifers are associated with older and denser economic centers: it may be the case that the type of people that sort into those places have more preference for high frequency amenities (they may be higher income, or with fewer children); at the same time, high density locations may be less conducive to commercial activity in low frequency sectors, perhaps because those sectors need larger retail surfaces. This might introduce an independent role for density on top of population size. We examine these possibilities on two sides.

In Table C.14 we extend our main regressions to control for an interaction between log income and log frequency of transactions. In this way, we control for the possibility that people with "tastes" for higher frequency sectors - as proxied by income - are more likely to sort in bigger places. In this table, the outcome variables are log employment (columns (1) and (2)), log establishment density (columns (3) and (4)), and log employees per establishment (columns (5) and (6)): within each of the outcomes, we replicate

Table C.11: **The role of individual heterogeneity (extended negative binomial regression)**

	(1)	(2)	(3)	(4)
Dependent Variable:	Number of transactions out of residence			
Number of transactions	0.051***	0.051***	0.043***	0.042***
	(0.001)	(0.001)	(0.001)	(0.001)
Log age	-0.061	-0.176	-0.103	
	(0.055)	(0.128)	(0.139)	
Log income	0.210***	0.483***	0.435***	
	(0.025)	(0.057)	(0.062)	
Log age × log frequency of transactions		-0.146***	-0.132**	-0.154***
		(0.053)	(0.054)	(0.054)
Log income × log frequency of transactions		-0.143***	-0.156***	-0.165***
		(0.023)	(0.022)	(0.022)
Log age × log of employees per store		0.113***	0.135***	0.145***
		(0.036)	(0.035)	(0.035)
Log income × log of employees per store		-0.039**	-0.055***	-0.051***
		(0.016)	(0.016)	(0.016)
Observations	28,959	28,959	28,959	28,959
Overdispersion	1.22	1.21	.69	.61
Sector fixed effects	No	Yes	Yes	Yes
ZIP code fixed effects	No	No	Yes	No
Individual fixed effects	No	No	No	Yes
Pseudo R-Square	.18	.18	.23	.24

Standard errors clustered at account level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

columns (5) and (6) in our main specification: columns differ on whether we control for heterogeneous state time trends or commuting zone time trends; all columns include sector-year fixed effects. As in our main specification, a higher income per capita is associated to a higher employment across sectors; moreover, employment tends to be smaller in high-frequency than in low frequency sectors as income grows. The sign of this correlation may reflect non-homothetic preferences and possibly capture further un-modeled equilibrium feedbacks. We do not see this exercise as a way to identify the causal effect of higher income, but rather as a simple way to control for heterogeneity in individual characteristics. Importantly, the sign and magnitude of the interaction between population and frequency is little affected.

In Table C.15, we directly examine the role of density, again on log employment (columns (1) and (2)), log establishment density (columns (3) and (4)), and log employees per establishment (columns (5) and (6)). The structure of the table mimics the one of Table C.14. The main regressors now include log income per capita, log density (computed as log population less log land area), an interaction of log density with

Table C.12: **The effect of rain (extended poisson regression)**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Number of transactions out of residence					
Number of transactions	0.023***	0.023***	0.023***	0.035***	0.036***	0.036***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Log age	-0.020	-0.019	-0.487***	-0.326**		
	(0.066)	(0.066)	(0.185)	(0.158)		
Log income	0.115***	0.115***	0.651***	0.408***		
	(0.032)	(0.032)	(0.089)	(0.066)		
Rain dummy	-0.314***	-0.954***	-0.948***	-0.978***	-0.980***	-1.292
	(0.019)	(0.055)	(0.054)	(0.048)	(0.048)	(0.881)
Rain dummy × log frequency of transactions		0.308***	0.308***	0.421***	0.424***	0.560
		(0.026)	(0.026)	(0.031)	(0.033)	(0.354)
Rain dummy × log of employees per store		0.047***	0.046***	0.024**	0.024**	0.089
		(0.013)	(0.012)	(0.009)	(0.010)	(0.205)
Log age × log frequency of transactions			-0.032	-0.053	-0.052	-0.088
			(0.069)	(0.058)	(0.058)	(0.068)
Log income × log frequency of transactions			-0.239***	-0.208***	-0.204***	-0.187***
			(0.037)	(0.025)	(0.025)	(0.028)
Log age × log of employees per store			0.195***	0.145***	0.145***	0.139***
			(0.048)	(0.032)	(0.032)	(0.038)
Log income × log of employees per store			-0.050**	-0.047***	-0.048***	-0.044***
			(0.022)	(0.014)	(0.014)	(0.015)
Rain dummy × log income						0.108*
						(0.062)
Rain dummy × log age						-0.243
						(0.166)
Rain × log income × log frequency of transactions						-0.047*
						(0.027)
Rain × log age × log frequency of transactions						0.105
						(0.072)
Rain × log income × log of employees per store						-0.011
						(0.014)
Rain × log age × log of employees per store						0.017
						(0.039)
Observations	57,918	57,918	57,918	57,918	57,918	57,918
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
ZIP code fixed effects	No	No	No	Yes	No	No
Individual fixed effects	No	No	No	No	Yes	Yes
Pseudo R-Square	.56	.56	.56	.68	.69	.69

Standard errors clustered at account level in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table C.13: **The effect of rain (extended negative binomial regression)**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Number of transactions out of residence					
Number of transactions	0.088***	0.090***	0.089***	0.072***	0.072***	0.072***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Log age	-0.075	-0.076	-0.257**	-0.244*		
	(0.054)	(0.054)	(0.120)	(0.130)		
Log income	0.194***	0.193***	0.500***	0.356***		
	(0.025)	(0.025)	(0.053)	(0.057)		
Rain dummy	-0.206***	-0.823***	-0.826***	-0.869***	-0.870***	-0.484
	(0.011)	(0.039)	(0.039)	(0.039)	(0.039)	(0.743)
Rain dummy × log frequency of transactions		0.493***	0.492***	0.466***	0.466***	0.542*
		(0.018)	(0.018)	(0.017)	(0.017)	(0.321)
Rain dummy × log of employees per store		-0.003	-0.002	0.004	0.003	-0.179
		(0.010)	(0.010)	(0.010)	(0.010)	(0.188)
Log age × log frequency of transactions			-0.115**	-0.119**	-0.124**	-0.167***
			(0.050)	(0.051)	(0.051)	(0.058)
Log income × log frequency of transactions			-0.149***	-0.163***	-0.166***	-0.149***
			(0.021)	(0.021)	(0.021)	(0.025)
Log age × log of employees per store			0.126***	0.141***	0.143***	0.092***
			(0.030)	(0.031)	(0.031)	(0.035)
Log income × log of employees per store			-0.045***	-0.055***	-0.054***	-0.045***
			(0.013)	(0.014)	(0.014)	(0.015)
Rain dummy × log income						0.122**
						(0.058)
Rain dummy × log age						-0.470***
						(0.139)
Rain × log income × log frequency of transactions						-0.042
						(0.026)
Rain × log age × log frequency of transactions						0.106*
						(0.060)
Rain × log income × log of employees per store						-0.021
						(0.015)
Rain × log age × log of employees per store						0.112***
						(0.036)
Observations	57,918	57,918	57,918	57,918	57,918	57,918
Overdispersion	1.12	1.1	1.09	.53	.52	.52
Sector fixed effects	No	No	Yes	Yes	Yes	Yes
ZIP code fixed effects	No	No	No	Yes	No	No
Individual fixed effects	No	No	No	No	Yes	Yes
Pseudo R-Square	.21	.21	.21	.27	.27	.27

Standard errors clustered at account level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table C.14: **Local outcome responses controlling for heterogeneity via income per capita**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log population	0.845***	1.247***	0.557***	0.811***	0.288***	0.436***
	(0.115)	(0.078)	(0.087)	(0.054)	(0.051)	(0.045)
Log population × log frequency	0.093***	0.082***	0.159***	0.153***	-0.065***	-0.070***
	(0.031)	(0.028)	(0.026)	(0.024)	(0.019)	(0.019)
Log land area	0.184***	-0.026	0.161***	0.033	0.023	-0.059**
	(0.034)	(0.044)	(0.026)	(0.029)	(0.015)	(0.026)
Log income per capita	2.576***	1.799***	2.201***	1.684***	0.375**	0.115
	(0.342)	(0.220)	(0.259)	(0.152)	(0.147)	(0.129)
Log income per capita × log frequency	-0.827***	-0.807***	-0.776***	-0.763***	-0.052	-0.044
	(0.085)	(0.077)	(0.074)	(0.068)	(0.051)	(0.051)
Sector-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State-Year Fixed Effects	Yes	No	Yes	No	Yes	No
Commuting Zone-Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-square	0.83	0.86	0.88	0.90	0.52	0.54
N	121,336	121,336	121,336	121,336	121,336	121,336
Cragg-Donald Wald F statistic	235.03	349.05	235.03	349.05	235.03	349.05
Kleibergen-Paap rk Wald F statistic	8.66	7.21	8.66	7.21	8.66	7.21

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

log frequency, and the main object of our analysis, log population times log frequency. We now have three endogenous variables: log density, its interaction with log frequency, and the interaction of population with frequency: as instruments, we use the county composition of consolidated and semi-consolidated aquifers, their interaction with frequency, and their interaction with county land size. These regressions are essentially "fixing" a density level and compare, say, a small sprawling place to a large sprawling place. We find that, controlling for density, an increase in population is significantly associated to a larger employment increase in high frequency sectors than in low frequency sectors. The point estimate of the interaction effect is much larger. The reason can be traced to the interaction of density and frequency. We find that as we increase density, employment in high frequency sectors *falls* relative to employment in low frequency sectors. This is consistent with our results in Figure 5 in the main text, that shows that adding denser counties to the estimation sample tends to pull the interaction coefficient down and towards zero.

Table C.15: **Local outcome responses controlling for density**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log density	0.321***	0.623***	0.199**	0.386***	0.121***	0.237***
	(0.107)	(0.084)	(0.083)	(0.060)	(0.036)	(0.037)
Log density × log frequency	-0.504***	-0.591***	-0.349***	-0.395***	-0.156***	-0.197***
	(0.077)	(0.072)	(0.059)	(0.053)	(0.025)	(0.025)
Log population × log frequency	1.106***	1.261***	0.845***	0.926***	0.261***	0.335***
	(0.063)	(0.043)	(0.049)	(0.031)	(0.027)	(0.027)
Log income per capita	1.803***	1.102***	1.485***	1.052***	0.318***	0.050
	(0.280)	(0.185)	(0.212)	(0.134)	(0.113)	(0.115)
Sector-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State-Year Fixed Effects	Yes	No	Yes	No	Yes	No
Commuting Zone-Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-square	0.76	0.77	0.81	0.83	0.50	0.51
N	121,336	121,336	121,336	121,336	121,336	121,336
Cragg-Donald Wald F statistic		236.22		236.22		236.22
Kleibergen-Paap rk Wald F statistic	4.97	4.53	4.97	4.53	4.97	4.53

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C.9 Commuting-adjusted population

This subsection describes further results on the differential impact of increases in commuting-adjusted population on local employment as a function of the frequency of transactions.

Table 11 below reports the effect of an increase in the commuting-adjusted population on the differential growth in the number of establishments across sectors; Table C.17 reports the same effect on the number of employees per establishment.

Table C.16: **Number of establishments and frequency of purchase: Commuting-Adjusted Population (C.a.p.)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log commuting-adjusted Population	0.548***	0.663***	0.603***	0.734***	0.639***
	(0.099)	(0.078)	(0.085)	(0.088)	(0.075)
Log commuting-adjusted population × log frequency	0.256***	0.213***	0.235***	0.201***	0.210***
	(0.077)	(0.070)	(0.072)	(0.064)	(0.066)
Log income per capita	1.061***	0.946***	1.013***	0.722***	0.917***
	(0.090)	(0.075)	(0.078)	(0.161)	(0.117)
Log land area	-0.049***	-0.036***	-0.041***	-0.056**	-0.011
	(0.008)	(0.008)	(0.007)	(0.026)	(0.027)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.88	0.89	0.89	0.90	0.90
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	84.43	84.22	169.16	113.12	168.14
Kleibergen-Paap rk Wald F statistic	11.92	11.76	13.11	2.83	4.03

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C.10 Likely Pure Consumers

This subsection describes further results on the differential impact of increases in likely pure consumers on local employment as a function of the frequency of transactions. Likely pure consumers are constructed by subtracting a measure of residential employment, the number of residents of a county working anywhere (either in the county or outside), from the total population of a county. The residential employment in a given year is constructed as follow. We start from the bilateral commuting matrices of 2007 (from the 5 year American Community Survey tabulations, 2005-2009), and 2000 (from the 2000 U.S. Census), cleaned using a procedure analogous to Monte et al. (2018). For each workplace county and year, we compute the origin composition of employment as the fraction of workers working in county i that reside in county n. We multiply this fraction for total employment in a county in 2007 and 1998, to obtain an estimate of the number of workers flowing from county n to county i. Summing across all destinations i for any given residence county n gives total residential employment.

Table 12 below reports the effect of an increase in the number of likely pure consumers on the differential growth in the number of establishments across sectors; Table C.19 reports the same effect on the number of employees per establishment.

Table C.17: **Number of employees per establishment and frequency of purchase: Commuting-Adjusted Population (C.a.p.)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log commuting-adjusted Population	0.352***	0.361***	0.355***	0.374***	0.388***
	(0.059)	(0.052)	(0.049)	(0.074)	(0.047)
Log commuting-adjusted population $\times$ log frequency	-0.057	-0.070*	-0.063*	-0.066*	-0.065*
	(0.039)	(0.040)	(0.034)	(0.035)	(0.034)
Log land area	-0.051***	-0.055***	-0.053***	-0.052**	-0.076***
	(0.008)	(0.008)	(0.008)	(0.026)	(0.025)
Log income per capita	0.029	0.083	0.058	0.052	0.066
	(0.076)	(0.074)	(0.071)	(0.161)	(0.105)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.51	0.54	0.53	0.53	0.55
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	84.43	84.22	169.16	113.12	168.14
Kleibergen-Paap rk Wald F statistic	11.92	11.76	13.11	2.83	4.03

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C.11 The role of fixed costs

In this section we control for the interaction between measures of fixed costs and frequency of transactions. A reasonable proxy is the economy-wide ratio of total employment to total establishments in a sector-year, i.e., the average establishment size: if fixed costs are high, increasing returns to scale tend to be important, and we should expect a higher employees-to-establishment ratio. In what follows, we will refer to this ratio simply as “fixed costs”.Table C.20 replicates the most conservative specifications in columns (5) and (6) for Tables 8-10: all the unreported columns behave similarly. Columns (1) and (2) replicate Table 8, where the dependent variable is employment. The coefficient on the interaction between frequency and population stays positive and of very similar magnitude. Moreover, in response to larger population, sectoral employment does not seem to change differentially as a function of fixed costs. The remaining 4 columns look at the margins of these changes: columns (3) and (4) replicate columns (5) and (6) of Table 9, where the dependent variable is the log number of establishments; columns (5) and (6) replicate the last two columns of Table 10, which consider employees per store. We find that controlling for fixed costs tilt these interactions slightly towards zero; the interaction with fixed costs are not estimated to be significant per se. Overall, we read these results as further evidence that consumers’ mobility impacts local economic outcomes.

Table C.18: **Number of establishments and frequency of purchase: Likely pure consumers (L.p.c.)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log likely pure consumers	0.425***	0.561***	0.491***	0.503***	0.750***
	(0.096)	(0.071)	(0.081)	(0.109)	(0.071)
Log likely pure consumers × log frequency	0.251***	0.223***	0.238***	0.215***	0.204***
	(0.053)	(0.049)	(0.050)	(0.047)	(0.044)
Log income per capita	1.703***	1.649***	1.687***	1.749***	1.333***
	(0.109)	(0.093)	(0.099)	(0.205)	(0.136)
Log land area	0.074***	0.101***	0.089***	0.160***	0.051
	(0.011)	(0.009)	(0.009)	(0.028)	(0.034)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.84	0.87	0.86	0.86	0.89
N	60,350	60,828	121,178	121,178	121,178
Cragg-Donald Wald F statistic	184.99	186.57	375.1	242.64	367.93
Kleibergen-Paap rk Wald F statistic	20.83	17.99	20.78	6.12	16.01

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C.12 Dependence on sector or areas

We examine whether our results are dependent on particular sectors being present in the regressions. In Table C.21, we report the results of running the most stringent specification in our main tables (column (6) in Tables 8-10) by removing one sector at a time: the first group of three columns report results for log employment, the second group for log establishments, and the third group for log employees per establishment. Within each group, we report the coefficient on the interaction between population and frequency, its standard error and its t-ratio. Regardless of the excluded sector, an increase in population increases employment in high frequency sectors more than in low frequency sectors; in no case this effect is insignificant; in one case, it is significant at 10% rather than 5%. In all cases, the increase occurs via more stores and a relatively smaller store sizes on average.

We further examine whether our results depend on particular geographical areas. In particular, we select randomly 5% of commuting zones and re-estimate the most stringent specification in our main tables (column (6) in Tables 8-10) without the counties in those commuting zones. We repeat this process 500 times. When the dependent variable is log employment, the interaction between population and frequency of transactions is positive and significant at 10% in three cases; in all the others, it is positive

Table C.19: **Number of employees per establishment and frequency of purchase: Likely pure consumers (L.p.c.)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log likely pure consumers	0.295***	0.323***	0.308***	0.293***	0.422***
	(0.045)	(0.040)	(0.039)	(0.056)	(0.048)
Log likely pure consumers × log frequency	-0.046	-0.063**	-0.054**	-0.056**	-0.061**
	(0.028)	(0.028)	(0.024)	(0.025)	(0.024)
Log land area	-0.005	-0.010	-0.008	0.019	-0.051**
	(0.006)	(0.006)	(0.006)	(0.016)	(0.025)
Log income per capita	0.262***	0.317***	0.291***	0.391***	0.229**
	(0.054)	(0.060)	(0.054)	(0.112)	(0.102)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.50	0.53	0.51	0.52	0.54
N	60,350	60,828	121,178	121,178	121,178
Cragg-Donald Wald F statistic	184.99	186.57	375.1	242.64	367.93
Kleibergen-Paap rk Wald F statistic	20.83	17.99	20.78	6.12	16.01

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

and significant at least at the 5% level. The estimated coefficients range between 0.055 and 0.142, with median 0.964. The distribution of coefficients for the three outcome variables, by level of significance, is reported in Figure C.4.

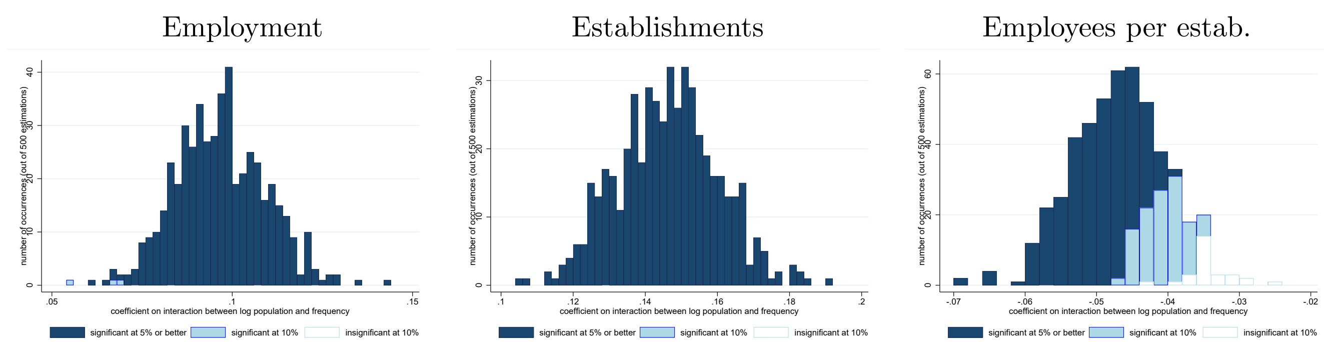

Figure C.4: Coefficients removing 5 percent of commuting zones

Table C.20: **Local outcome responses controlling for fixed costs**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log population	0.709***	1.120***	0.493***	0.758***	0.216**	0.362***
	(0.128)	(0.117)	(0.132)	(0.099)	(0.085)	(0.100)
Log population × log frequency	0.113***	0.097**	0.151***	0.141***	-0.038*	-0.045**
	(0.042)	(0.038)	(0.034)	(0.031)	(0.022)	(0.021)
Log population × log fixed costs	0.044	0.042	0.026	0.023	0.017	0.019
	(0.030)	(0.030)	(0.024)	(0.024)	(0.033)	(0.034)
Log land area	0.183***	-0.024	0.161***	0.036	0.023	-0.060**
	(0.034)	(0.043)	(0.026)	(0.029)	(0.015)	(0.026)
Log income per capita	1.765***	1.012***	1.440***	0.943***	0.324***	0.069
	(0.294)	(0.202)	(0.216)	(0.133)	(0.125)	(0.118)
Sector-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State-Year Fixed Effects	Yes	No	Yes	No	Yes	No
Commuting Zone-Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-square	0.83	0.85	0.87	0.90	0.53	0.54
N	121,336	121,336	121,336	121,336	121,336	121,336
Cragg-Donald Wald F statistic	155.19	230.38	155.19	230.38	155.19	230.38
Kleibergen-Paap rk Wald F statistic	3.86	4.6	3.86	4.6	3.86	4.6

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C.21: **Robustness to excluding one sector at a time**

Excluded Sector	Employment			Establishments			Employees per establishment		
	(coeff)	(s.e.)	(t ratio)	(coeff)	(s.e.)	(t ratio)	(coeff)	(s.e.)	(t ratio)
Agricultural Services	0.09	0.04	2.36	0.14	0.03	4.20	-0.05	0.02	2.16
Amusement, Rec. Serv.	0.10	0.04	2.74	0.15	0.03	4.57	-0.05	0.02	2.25
Apparel	0.10	0.04	2.54	0.14	0.03	4.45	-0.05	0.02	2.28
Auto Repair/Service/Parking	0.10	0.04	2.49	0.15	0.03	4.53	-0.05	0.02	2.41
Auto and Truck Sales/Service/Parts	0.10	0.04	2.60	0.15	0.03	4.37	-0.04	0.02	2.05
Building Mat./Hardware/Garden Supp.	0.13	0.04	3.24	0.18	0.04	5.04	-0.05	0.02	2.27
Communications	0.10	0.04	2.79	0.15	0.03	4.37	-0.04	0.02	2.14
Durable Goods	0.08	0.04	2.17	0.15	0.03	4.65	-0.07	0.02	3.16
Eating and Drinking Places	0.09	0.04	2.19	0.14	0.04	4.06	-0.06	0.02	2.61
Food Stores	0.11	0.04	2.78	0.12	0.03	4.26	-0.01	0.02	0.42
Furniture, Home Furnishings, Equip.	0.10	0.04	2.66	0.15	0.03	4.54	-0.04	0.02	2.08
Gasoline Services	0.15	0.04	3.43	0.16	0.04	4.52	-0.02	0.03	0.64
General Merchandise Stores	0.07	0.04	1.92	0.13	0.03	4.87	-0.06	0.02	3.03
Health Services	0.10	0.04	2.53	0.16	0.03	4.73	-0.06	0.02	2.63
Hospitality	0.08	0.04	2.36	0.13	0.03	4.29	-0.05	0.02	2.36
Misc. Retail	0.07	0.04	1.97	0.13	0.03	4.02	-0.06	0.02	2.61
Misc. Services	0.14	0.04	3.12	0.18	0.04	4.97	-0.04	0.02	1.85
Motion Pictures	0.10	0.04	2.56	0.14	0.03	4.40	-0.05	0.02	2.11
NonDurable Goods	0.07	0.04	1.76	0.11	0.03	3.72	-0.05	0.02	2.11
Other Vehicles Sales/Service/Parts	0.10	0.04	2.55	0.14	0.03	4.35	-0.04	0.02	1.95
Personal Services	0.10	0.04	2.64	0.15	0.03	4.53	-0.05	0.02	2.18

C.13 Weak Instruments

As an alternative check on the strength of our instrumentation strategy, we estimate the main tables in our analysis via Limited Information Maximum Likelihood (LIML), rather than Two-Stages Least Squares (2SLS). LIML estimators are known to have better small sample properties with weak instruments. Swings in the coefficients or much larger standard errors as compared to 2SLS would be an indication of a potentially weak instrument.

Table C.22-C.24 report the corresponding LIML estimates of Tables 8-10. Coefficients are broadly in line, and standard errors essentially unchanged. These findings further help alleviate potential concerns about weak instruments.

Table C.22: **Local employment and frequency of purchase: LIML estimator**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log employment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.749***	0.902***	0.826***	0.812***	1.223***
	(0.103)	(0.072)	(0.083)	(0.126)	(0.085)
Log population × log frequency	0.155***	0.105**	0.129***	0.117***	0.100***
	(0.052)	(0.043)	(0.044)	(0.043)	(0.039)
Log income per capita	1.641***	1.604***	1.632***	1.783***	1.016***
	(0.149)	(0.137)	(0.137)	(0.303)	(0.205)
Log land area	0.103***	0.136***	0.120***	0.185***	-0.024
	(0.013)	(0.011)	(0.011)	(0.035)	(0.044)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.81	0.84	0.83	0.83	0.85
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	205.05	223.19	435.95	233.1	346.08
Kleibergen-Paap rk Wald F statistic	23.42	21.73	24.93	5.68	6.15

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We also use the full set of 6 types of aquifers (i.e, including sandstone aquifers, carbonated-rock aquifers, sandstone and carbonate-rock aquifers, igneous and metamorphic-rock aquifers) rather than just consolidated and semiconsolidated aquifers as instruments for population in Table 8-10. Tables C.25-C.27 report the results. The Cragg-Donald statistics become a little weaker and the coefficients become smaller

Table C.23: **Number of establishments and frequency of purchase: LIML estimator**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.441***	0.591***	0.520***	0.528***	0.797***
	(0.096)	(0.061)	(0.074)	(0.106)	(0.065)
Log population × log frequency	0.206***	0.163***	0.183***	0.165***	0.157***
	(0.049)	(0.039)	(0.042)	(0.039)	(0.036)
Log income per capita	1.498***	1.385***	1.446***	1.499***	0.953***
	(0.135)	(0.104)	(0.114)	(0.246)	(0.145)
Log land area	0.098***	0.135***	0.118***	0.166***	0.038
	(0.012)	(0.008)	(0.009)	(0.029)	(0.031)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.85	0.88	0.86	0.87	0.90
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	205.05	223.19	435.95	233.1	346.08
Kleibergen-Paap rk Wald F statistic	23.42	21.73	24.93	5.68	6.15

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

and towards the OLS result; the patterns of significance are broadly preserved.

Table C.24: **Number of employees per establishment and frequency of purchase: LIML estimator**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.281***	0.299***	0.288***	0.259***	0.411***
	(0.042)	(0.036)	(0.035)	(0.055)	(0.047)
Log population × log frequency	-0.038	-0.048*	-0.043**	-0.040*	-0.047**
	(0.025)	(0.025)	(0.022)	(0.022)	(0.021)
Log land area	0.004	0.002	0.002	0.024	-0.059**
	(0.006)	(0.006)	(0.006)	(0.015)	(0.026)
Log income per capita	0.174***	0.229***	0.204***	0.335***	0.076
	(0.059)	(0.064)	(0.058)	(0.130)	(0.118)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.51	0.53	0.52	0.52	0.54
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	205.05	223.19	435.95	233.1	346.08
Kleibergen-Paap rk Wald F statistic	23.42	21.73	24.93	5.68	6.15

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C.25: **Local employment and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			county-sector log employment		
Sample years :	07	98	98,07	98,07	98,07
Log population	0.975***	1.017***	0.985***	0.981***	1.226***
	(0.045)	(0.043)	(0.043)	(0.066)	(0.060)
Log population × log frequency	0.057**	0.037	0.048**	0.046**	0.042*
	(0.025)	(0.024)	(0.023)	(0.023)	(0.022)
Log income per capita	1.347***	1.463***	1.423***	1.494***	1.165***
	(0.077)	(0.094)	(0.083)	(0.170)	(0.153)
Log land area	0.111***	0.135***	0.122***	0.157***	0.008
	(0.010)	(0.010)	(0.010)	(0.023)	(0.034)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.84	0.84	0.84	0.84	0.86
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	168.11	156.3	323.75	175.28	192.93
Kleibergen-Paap rk Wald F statistic	13.38	14.67	14.75	6	4.76

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C.26: **Number of establishments and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.676***	0.684***	0.669***	0.690***	0.844***
	(0.034)	(0.033)	(0.033)	(0.048)	(0.041)
Log population $\times$ log frequency	0.097***	0.093***	0.097***	0.089***	0.088***
	(0.018)	(0.019)	(0.018)	(0.017)	(0.017)
Log income per capita	1.208***	1.315***	1.278***	1.242***	1.013***
	(0.059)	(0.070)	(0.063)	(0.123)	(0.103)
Log land area	0.106***	0.134***	0.119***	0.142***	0.051**
	(0.008)	(0.008)	(0.008)	(0.017)	(0.023)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.88	0.88	0.88	0.89	0.90
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	168.11	156.3	323.75	175.28	192.93
Kleibergen-Paap rk Wald F statistic	13.38	14.67	14.75	6	4.76

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C.27: **Number of employees per establishment and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.299***	0.333***	0.316***	0.290***	0.382***
	(0.024)	(0.024)	(0.022)	(0.033)	(0.034)
Log population × log frequency	-0.040**	-0.056***	-0.048***	-0.043***	-0.046***
	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)
Log land area	0.005	0.001	0.003	0.016	-0.043**
	(0.006)	(0.006)	(0.006)	(0.012)	(0.019)
Log income per capita	0.139***	0.148***	0.145***	0.252***	0.152*
	(0.039)	(0.050)	(0.042)	(0.083)	(0.087)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.51	0.53	0.52	0.53	0.54
N	60,413	60,923	121,336	121,336	121,336
Cragg-Donald Wald F statistic	168.11	156.3	323.75	175.28	192.93
Kleibergen-Paap rk Wald F statistic	13.38	14.67	14.75	6	4.76

Standard errors clustered at county level in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$