The Geography of Consumption*

Sumit Agarwal† J. Bradford Jensen†‡ Ferdinando Monte†§

November 11, 2017

(please click here for the latest version)

Abstract

We use detailed information from U.S. consumers' credit card purchases to provide the first large-scale description of the geography of consumption. We find that consumers' mobility is quite limited and document significant heterogeneity in the importance of gravity across sectors. Gravity is stronger in more frequently purchased sectors. Consumers actively manage the spatial dimension of their consumption choices: using daily rain precipitation from thousands of weather stations in U.S., we show that shocks to travel costs change the spatial distribution of expenditure, and they do so differentially across sectors. These choices matter for local equilibrium outcomes: using underlying geological variation across U.S. counties, we show that sectors with high storage costs respond with larger employment and denser stores to exogenous differences in population. This response is consistent with a model where consumers optimally choose to travel more frequently and for shorter distances for their purchases in higher storage costs sectors. Our results suggest that incorporating the demand-side is necessary to analyze the distributional consequences of local and aggregate shocks across regions. These results also suggest the demand-side is critical to understanding the location of firms and employment in the large and understudied service sector.

JEL codes: R1, R2, F1, F14, L8

*Sumit Agarwal: sa1106@georgetown.edu; Brad Jensen: jbj24@georgetown.edu; Ferdinando Monte: ferdinando.monte@georgetown.edu. We would like to thank Steve Redding, Andrés Rodriguez-Clare, Esteban Rossi-Hansberg, and Tony Yezer for helpful comments and discussions.

†Georgetown University, McDonough School of Business, 37th and O streets, NW, Washington DC 20057.

‡NBER

§CEPR

1 Introduction

The equilibrium production and location decisions of firms depends, among other things, on the characteristics of local demand and, in particular, on the willingness of consumers to travel to buy goods and services. However, lack of direct evidence on consumers' mobility has limited our ability to characterize local consumption markets. In this paper, we focus on the consumers' demand side and provide the first large scale description of the geographical dimension of different consumption markets using more than 1.7 million individual American consumers' credit card transactions. We show that consumers actively manage the spatial dimension of their purchases and that this has measurable consequences on local equilibrium outcomes.

Our data reveals that consumer mobility is very limited: agents typically source their purchases from just a few of the many locations available to them. Moreover, as broadly documented for merchandise trade both at international and intra-national level, gravity is a first-order feature of the data. Comparing at–home and out–of–home behavior, we find that large drops in expenditure occur already 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 that gravity is present but milder with respect to 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 intra-national trade (-1.3%) or commuting flows (-4.4%).[1]

To understand the sources of this decline, we decompose total expenditure in contributions coming from 1) 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 transaction (a "batch size" intensive margin). We analyze the spatial decay of each component separately and by sector. In all cases, extensive margins account for almost the totality of the decline. We find large heterogeneity across industries in the overall impact of distance, which largely reflects the importance of the frequency margin. Moreover, we find that stronger gravity in an industry is correlated with a higher average frequency of transactions in the data.

Since the cost of a trip is largely independent from the volume purchased, heterogeneity in the durability or storability of goods and services can produce heterogeneous spatial purchasing patterns across sectors: when goods are less storable, consumers will want to buy smaller batches, more frequently, but closer to home. We think of storability/durability as a general characteristic of the sector, capturing the length of time by which a good or service can deliver its utility flow. For perishable items this concept is intuitive: fruits for example will depreciate if not eaten soon. For durable goods, this concept may reflect the 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 them 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 also do. While economists have long studied how exogenous characteristics of an industry

[1] See for example Hillberry and Hummels (2007), and Monte, Redding and Rossi-Hansberg (2017).

(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, and 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 related to the durability/storability of the good. In what follows, we will refer to "storability" and "storage costs" for brevity, but have this more general interpretation in mind.

To ensure that our cross–sectional findings are not driven by unobserved heterogeneity on the supply side, we use exogenous variation in consumers' travel costs induced by rainfall. We use daily data on rain precipitation from thousands of weather stations to examine the response of expenditure to higher travel costs over time within origin–destination pairs: by shocking travel costs over time while supply is fixed, we can test whether the spatial dimension of travel is an active choice margin, and whether such margin is heterogeneous across sectors.

We find that consumers actively manage the spatial dimension of their demand. Rain reduces expenditure of consumers living in a particular place both at home and in outside locations. Gravity, however, becomes flatter, and the common rain shock affects gravity heterogeneously across sectors: in sectors more frequently purchased, gravity is affected the least. These results are consistent with rain inducing consumers most sensitive to travel costs not to go out and spend, so that only those less sensitive to travel costs are observed purchasing. Moreover, the fact that frequently purchased goods are the least affected suggests that those are also the ones with higher storage costs, where average inventories are the lowest.

These considerations suggest that from the perspective of a merchant, the consumer's travel behavior is shaped by the characteristics of storability or durability of the goods and services. However, rain shocks are not apt at identifying the causal effect of changes in travel costs on overall expenditure and other local outcomes. We then examine whether the spatial distribution of economic activity is affected by consumers' spatial choices in cross–sections of county data: in response to an increase in local population, demand should be more geographically concentrated for sectors with higher storage costs, and hence supply should respond by increasing local employment faster in those sectors. Moreover, since the differential concentration of demand is driven by the desire of consumers to save time, the additional employment should come about in the form of more establishments (thereby reducing the average distance between consumers and stores) rather than more employees per establishment.

We test this prediction using two measures of storage costs: we first use the average observed frequency of transactions as a simple proxy, and then we recover an index of storage costs from a simple model where consumers optimally choose how frequently and how far to travel to procure their goods and services. We use the underlying geological composition of a county to circumvent the endogeneity problem arising from regressing employment in a county–sector on population in the county itself. 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 addition of establishments at a faster rate (i.e., increase in density), while employment per establishment grows at a slower rate. These findings are consistent with a more geographically concentrated demand arising from the need to save on travel time.

Our analysis shows that the structure of the purchasing technology available to optimizing consumers (cost of a trip increasing in distance but fixed with respect to transaction size) interacts with heterogeneity in storage costs to generate measurable differences in local outcomes across sectors. These effects are potentially very pervasive: final consumption accounted for around 70% of GDP in 2015 in the United States; the service industries involved in its delivery, from apparel stores, to restaurants and personal services providers, accounted for around 70% of employment, and more than 80% of total value added. Understanding the nature of spatial patterns of consumption is therefore essential for a wide spectrum of issues: from the degree of spatial competition between firms to the determination of local and aggregate productivity and factors' income, from the consequences of local labor demand shocks, local taxes and regulation, to the impact of investment in transportation infrastructure and of other "place-based policies". Further, our results provide important information for the study of the liberalization of international trade and investment in services, since the extent to which foreign direct investment flows to a country (and to which part of the country) depends, among other things, on how local the market for a particular service is. In this sense, our analysis relate to Alessandria, Kaboski and Midrigan (2010), which shows that the microstructure of firms' transaction technology has aggregate consequences on the level of trade, and regulates the response of prices and import volumes to large devaluations.

While spatial analyses of the manufacturing sector abound, they are not particularly informative regarding the final consumer behavior. The practical importance of direct sales from manufacturers to consumers is still somewhat limited: for example, e-commerce sales (which include sales from a company's website, but also indirect sales from other distributors) account for only 6.4% of total retail sales in 2014, and only 0.9% in 2000, closer to our sample period, 2003 (Hortaçsu and Syverson, 2015). Intra-national surveys on goods' flows typically record firm-to-firm transactions. The limited literature on spatial competition in consumption markets mostly focuses on specific industries. For the restaurant industry, Couture (2016) evaluates the extent to which consumers gain from increased density and finds that larger variety (rather than lower travel times) are the main driver of consumers' valuation of higher density; Davis, Dingel, Monras and Morales (2017) discuss the relative impact of spatial vs. social friction in restaurant consumption. In the food distribution sector, Handbury, Rahkovsky and Schnell (2016) study the role of spatial access to healthy food supply in explaining differences in the quality of food intake across income groups, and argue that observed differences in access are most plausibly the result of optimal supply responses to differences in demand. Other industries that have been studied include gasoline (Houde, 2012) and movie theaters (Davis 2006). Studies focusing on consumption across cities, with less focus on consumers' mobility, include Glaeser, Kolko and Saiz (2000) (who explore the increasing importance of cities as consumption centers) and Schiff (2015), who finds that larger and denser cities offer more restaurant varieties. Overall, we contribute to this literature by providing results which are comparable across industries, extend the set of industries for which we can assess consumers' mobility, and exploit cross-industry variation to argue that storability of a good is effectively a determinant of gravity. We do so by building on and extending the literature on spatial frictions (Anderson 1979, Anderson and Van Wincoop 2003, Eaton and Kortum 2003, Hummels and Klenow 2005, Hillberry and Hummels 2008).

Our results are also related to the literature on cross–border consumption behavior. For example, Chandra, Head and Tappata (2014) find that consumers' trip counts across the U.S.–Canada border responds to real exchange rate movements and distance to the border; they propose and estimate a simple model to rationalize those findings. Agarwal, Marwell and McGranahan (2017) show that consumers cross state lines in response to state tax holidays and permanently reallocate expenditure away from other goods. We generalize these findings on consumers' mobility and emphasize travel and heterogeneous storage costs: this focus allows us to study the nature of cross-sectoral differences in the strength of gravity and its relation with the frequency of transactions.

The retail sector is also subject of a growing literature. In an international context, Bernard, Jensen, Redding and Schott (2010) examine characteristics of wholesalers and retailers involved in international transactions, finding that they are significantly smaller compared to their "producer and consumer" counterparts. Jarmin, Klimek and Miranda (2005) and Hortaçsu and Syverson (2015) present some overall trends in the industry. The important role of the retail sector in price determination is emphasized by Nakamura (2008), who shows that a majority of retail-store price variation is attributable to retail chain–level shocks, while only a minor fraction to manufacturers or wholesalers–specific shocks. These findings emphasize the importance of understanding pricing (and hence demand) conditions at the retail level: while our data is lacking pricing information, we are contributing to the understanding of the nature of demand.

Our work is also relevant for the growing literature on e-commerce and on-line transactions: firms operating in this way are competing with brick-and-mortar stores precisely taking into account consumers' travel costs. Aspects of on-line vs. off-line retail are analyzed in Ellison and Ellison (2009), who study the importance of taxes in determining sales of on-line versus traditional retailers: among other things, they find that geography still matters (consumers prefer to buy from home state or neighboring retailers after accounting for other factors), albeit the effect of proximity via shipping times is small. The importance of distance and the persistence of a home bias is also found in on-line auctions (Hortaçsu, Martínez-Jerez and Douglas 2009). Einav et al. (2017) use credit card transactions to quantify the gains from e-commerce.

We proceed in the paper by presenting some cross–sectional evidence in Section 2. In Section 3 we turn to a within-sector analysis and show how the spatial distribution of expenditure responds to rainfall shocks. Section 4 shows that consumer behavior matters for local economic outcomes. Section 5 discusses some further implications of our results, robustness and limitations of our results. Section 6 concludes.

2 The Geography of Consumption

2.1 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 February and October 2003. A transaction record contains, among other things, an exact date, an account ID, the amount spent, a Merchant Category Code (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; importantly, we know the ZIP code associated to it. The same data has been used in Agarwal, Marwell, and McGranahan (2017). After cleaning procedures, we have 1,751,067 transactions for 71,927 accounts (see Appendix A for a complete description of the data cleaning and processing). The average transaction is 70 dollars, and total purchases amount to around $122 million. Table 1 reports a breakdown by 24 broad categories. The largest categories in terms of observations are Gasoline Services, Food Stores, Miscellaneous Retail, and Eating and Drinking Places. Table B.1 in the Appendix (page 41) shows summary statistics by State of purchase. The largest number of transactions are reported in New York, California, and Massachusetts.

In the remainder of this section we establish a number of stylized facts on the local nature of consumption markets. This analysis will be inherently cross-sectional in nature. In the following section we will study the consequences of a cost shock to show that the spatial distribution of expenditure reacts over time to consumers' travel costs for a fixed location of consumers and merchants.

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

Broad Category	**Median**	**Mean**	**St. Dev.**	**Sum**	**N**
Agricultural Services	83	136	212	1,307,704	9,616
Amusement, Rec. Serv.	45	89	169	1,772,753	19,912
Apparel	49	75	114	6,117,925	81,846
Auto Repair/Service/Parking	41	151	325	3,465,225	23,009
Auto and Truck Sales/Service/Parts	66	198	423	6,625,779	33,486
Building Mat./Hardware/Garden Supp.	42	101	258	9,662,827	95,641
Communications	53	91	122	559,309	6,115
Durable Goods	68	209	520	839,482	4,026
Eating and Drinking Places	26	39	73	8,785,594	228,006
Educational Services	92	298	625	2,296,223	7,711
Food Stores	30	46	59	12,126,085	266,030
Furniture, Home Furnishings, Equip.	60	194	430	10,865,485	55,999
Gasoline Services	19	22	31	6,938,938	312,873
General Merchandise Stores	43	67	122	13,971,247	207,985
Health Services	71	164	375	4,490,768	27,403
Hospitality	96	170	308	6,436,342	37,971
Misc. Retail	32	65	182	16,113,793	248,288
Misc. Services	95	316	702	1,873,105	5,927
Motion Pictures	14	19	44	273,520	14,080
NonDurable Goods	38	78	175	640,316	8,249
Other Vehicles Sales/Service/Parts	76	259	746	1,366,524	5,280
Personal Services	37	74	212	2,417,176	32,575
Transportation Services	38	117	384	1,128,331	9,621
Vehicle Rental	125	189	230	1,778,716	9,418
Total	30	70	195	121,853,167	1,751,067

2.2 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. The data identifies expenditure flows between 17,572 unique locations (11,474 unique residence and 14,997 unique sale locations). The data records transactions among 238,269 unique pairs, 0.077% of all possible pairs. There are 7.3 transactions per pair, and the median pair has 1 transaction. Naturally, it is unrealistic to expect consumers to travel very long distances: for example, 2,588,158 potential location pairs (about 0.83% of all pairs) in our data have distance below 120 km.[2] Among those pairs, expenditure flows are recorded between only 122,130 pairs, or 4.7% of the possible pairs, with 14.3 transactions per pair on average, and 2 transactions for the median pair. Overall, the matrix of residence-sales location purchases is sparse.

Table 2 digs deeper into this limited consumer mobility. Its first row shows that consumers in the median residence visit only 6 distinct sales locations overall during the sample period (10.6 sales locations on average). One might think that this low number is simply a consequence of absence of close-by options. This is not the case. The second row in the Table shows that consumers living in the median residence also have 186 sales locations within 120 km; the third row shows that overall the fraction of available locations where purchases actually occur is very small.[3]

Table 2: **Summary statistics across residence locations**

variable	**min**	**p10**	**p25**	**p50**	**p75**	**p90**	**max**	**mean**	**N**
Sales locations visited	1	1	3	6	13	26	447	10.92	11,179
Sales locations within 120km	2	61	109	186	327	636	1,116	265.73	11,179
Share available loc. visited	0	0.01	0.02	0.04	0.07	0.13	0.68	0.06	11,179

What accounts for this low mobility? Overall, the number of visited locations grows at about half the pace of the available locations. In column 1 of Table 3 we regress the log of number of sales locations visited on the log number of sales locations available and find an elasticity of 0.54 (hence, well below 1). 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 2.5% decrease in the number of locations visited (column 3). Such a high elasticity suggests a central role of distance on consumers' expenditure patterns. We explore this aspect next.

[2]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 (2017) find this threshold to be one where gravity in home-to-work commuting flows has a structural break, and seems to be a natural cutoff to focus on.

[3]In Appendix B.2, page 40, we further show that this low mobility is not driven by accounts with low credit card usage overall.

Table 3: Locations available and locations visited

Dependent variable:	Log of number of sales locations visited (1)	(2)	(3)
Sales locations within 120km, log	0.539***		0.561***
	(0.010)		(0.009)
Average distance to sales locations within 120km, log		-2.361***	-2.594***
		(0.074)	(0.064)
Constant	-0.957***	12.011***	10.097***
	(0.051)	(0.319)	(0.276)
R^2	0.22	0.08	0.32
N	11,179	11,179	11,179

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

2.3 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; furthermore, a strong role is played by distance. A natural question to ask then is: how much do people travel for their purchases?

The median transaction in the data occurs at 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.[4] 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.[5] 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.[6] 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.

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

[5] Tables B.3 and B.4, in pages 43 and 44 respectively, show percentiles in the distribution of transaction distances by sector in the raw data and weighted by value of the transaction, respectively. The typical dollar is spent farther than where the typical transaction occurs, as reflected in right-ward shifts in the value-weighted distributions.

[6] We will show below a more precise relation between the importance of distance and the frequency of transactions.

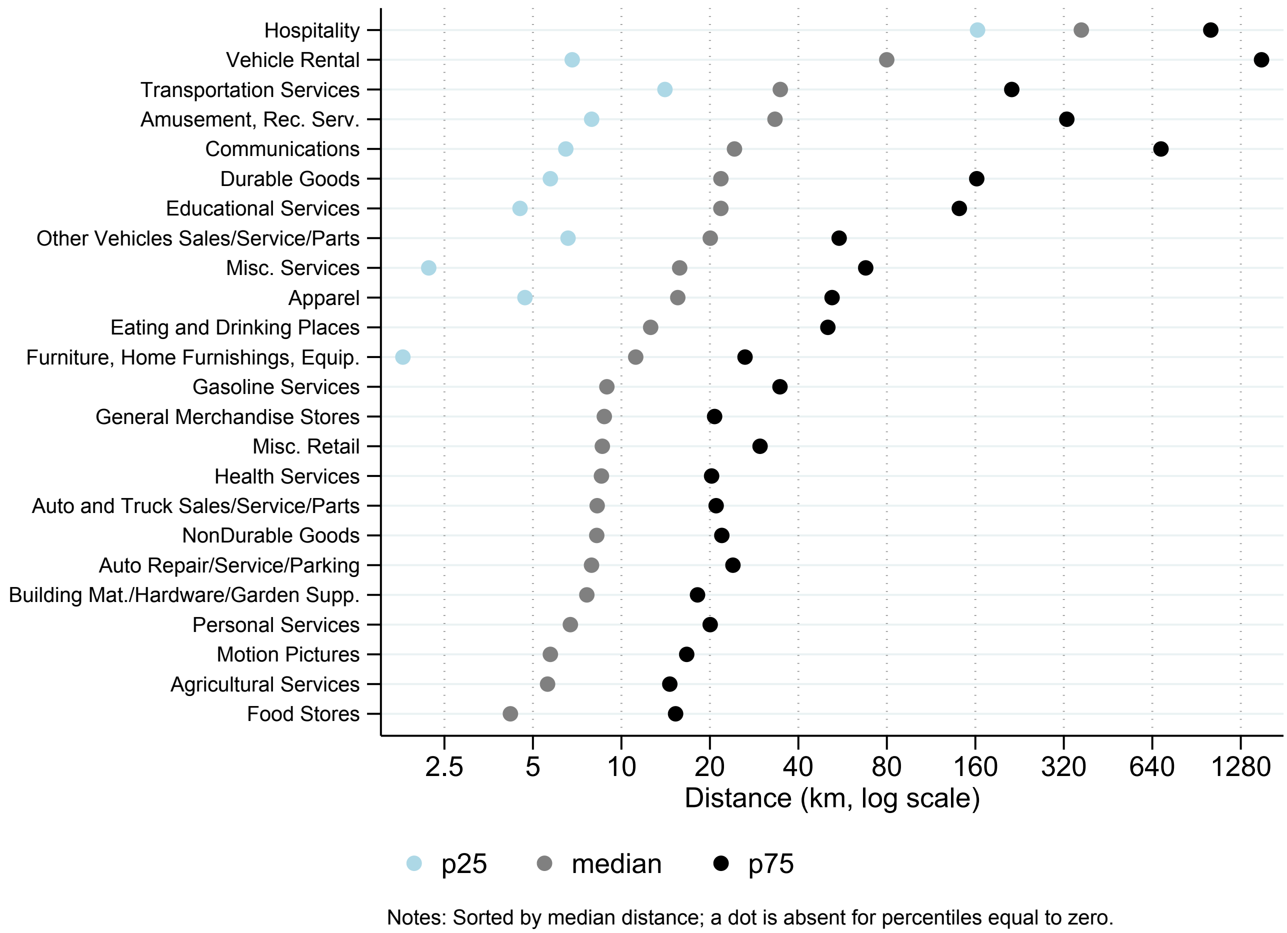

Figure 1: Distances traveled by sector (select percentiles)

Obviously, the measured 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 consumers' 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.

2.4 Gravity in Consumer Expenditure

Gravity is an almost universal feature of spatial relationships. While a large literature has documented the decay of goods' trade flows with distance at inter-national and intra-national level, little is known about the spatial behavior of consumers.[7] We fill this gap in two steps.

First, we document that gravity also holds for consumers' behavior[8]. 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 home.

We then analyze the determinants 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).

2.4.1 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 only in a minority of cases, our knowledge of gravity in final consumption is extremely limited. Moreover, virtually all the literature deals with merchandise shipments, thus ignoring the service sector altogether. Here, we fill these important gaps.

Denote with $x_{a,m}$ the observed expenditure of account a falling on merchant m in a sector in the whole sample period. Note that each account a has an associated home location $h = r(a)$, and each merchant has a sale location $s = l(m)$. We start by aggregating our expenditure at the sector – account home – sales location level, X_{hs}.

$$X_{hs} \equiv \sum_{a:r(a)=h,m:l(m)=s} x_{a,m}$$

We initially relate the expenditure X_{hs} to the distance between residence and merchant's shop in two ways. First, we simply estimate the change in expenditure associated with shopping out of the home residence:

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

where $\mathbf{1}_{(h \neq s)}$ is an indicator function assuming the value of 1 if $h \neq s$ and zero otherwise. Second, we follow the gravity literature and estimate the impact of distance on trade flows with a regression of the form

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

including only pairs where $h \neq s$; in this equation, α is a constant, and $dist_{hs}$ is the distance between the centroids of h and s. 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.

[7] International flows of goods are only measured at country level, thus ignoring the travel dimension of consumers' purchase. Intra-national flows of goods typically record firm-to-firm transactions.

[8] To our knowledge, the first formulation of a gravity law in the retail sector dates back to 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."

Eq. (1) measures the drop in expenditure associated by visiting the average location out of home, and hence it shows the importance of very short trips, for which, however, distance is poorly measured. Eq. (2) shows the elasticity of expenditure to distance comparing pairs of locations outside home, in which case distance can be measured.

We first estimate equations (1) and (2) across all sectors, using distances up to 120 km. We find, unsurprisingly, very clear effects of distances. Estimating (1), the average expenditure out of home is about 8.8% of the average expenditure at home ($\eta = -2.428$, robust s.e. 0.021).[9] When we estimate (2), we find a slope of -1.049 (s.e. of 0.006), in line with estimates in the trade literature.[10] A comparison of these two coefficients shows that a large decay appears to occur already 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 eq. (1) and (2) by sector. 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, such decay is heterogeneous: in sectors like Food Stores, the point estimate of average expenditure out of home is around 10% 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 eq. (2) is consistent with this picture: the correlation between the two sets of coefficients across sectors is 0.68.

2.4.2 Margins of adjustment

Why does expenditure decay with space? One can think that as distance increases, there may be fewer people traveling out; moreover, those whom 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

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

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

Eq. (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, 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), as emphasized in (4).

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

[10] This slope is not particularly sensitive to changes in the cutoff. See Appendix B.4, page 40, for further discussion.

Table 4: **Decline in expenditure**

Category	Out of Home coeff (1)	Out of Home pv (2)	Out of Home obs. (3)	Gravity coeff (4)	Gravity pv (5)	Gravity obs. (6)
Food Stores	-2.23	0.00	22,652	-0.85	0.00	18,635
Gasoline Services	-2.08	0.00	39,673	-0.60	0.00	34,621
General Merchandise Stores	-1.79	0.00	26,845	-0.93	0.00	23,933
Misc. Retail	-1.70	0.00	34,057	-0.65	0.00	30,046
Eating and Drinking Places	-1.57	0.00	34,509	-0.56	0.00	31,028
Building Mat./Hardware/Garden Supp.	-1.40	0.00	14,190	-0.73	0.00	11,610
Auto Repair/Service/Parking	-1.25	0.00	4,415	-0.40	0.00	3,014
NonDurable Goods	-1.16	0.00	978	-0.65	0.00	758
Health Services	-1.12	0.00	5,136	-0.33	0.00	3,914
Apparel	-1.10	0.00	15,921	-0.53	0.00	14,069
Transportation Services	-1.09	0.00	743	-0.47	0.00	635
Furniture, Home Furnishings, Equip.	-1.07	0.00	12,292	-0.57	0.00	10,740
Auto and Truck Sales/Service/Parts	-1.04	0.00	7,302	-0.33	0.00	5,508
Motion Pictures	-1.04	0.00	1,927	-0.34	0.00	1,253
Amusement, Rec. Serv.	-1.02	0.00	2,959	-0.22	0.00	2,330
Educational Services	-1.00	0.00	712	-0.15	0.38	530
Personal Services	-0.96	0.00	5,204	-0.31	0.00	3,761
Vehicle Rental	-0.95	0.00	546	-0.08	0.59	296
Misc. Services	-0.92	0.06	222	0.97	0.01	120
Communications	-0.89	0.00	424	-0.41	0.01	263
Agricultural Services	-0.88	0.00	552	0.42	0.11	190
Other Vehicles Sales/Service/Parts	-0.68	0.41	257	-0.59	0.08	128
Hospitality	-0.65	0.01	1,394	-0.14	0.08	1,160
Durable Goods	-0.09	0.90	79	1.11	0.67	15

When we re-estimate eq. (1) with the left hand-side being each of these three terms, the coefficients on the out-of-home dummy add up to the overall coefficients η reported in column 1 of Table 4 (and similarly for eq. (2)).[11]

Figure 2 shows the results of this decomposition for eq. (1). The length of each bar corresponds to column 1 in Table 4. 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.[12] As a benchmark, Hillberry

[11] A further angle of this decomposition could relate to the Allen and Alchian (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 mesurement of unit values and hence cannot be used to speak to this conjecture. For related work, see Hummels and Skiba (2004).

[12] Figure B.2 in the Appendix (p. 47) shows the same decomposition for eq. (2): the extensive margin is even less important, accounting for 57% of the decline in expenditure in the typical sector as distance increases. Tables B.5 and B.6, also in the Appendix (pages 45), show the actual values of the account and expenditure margin with associated p-values.

and Hummels (2007) find, for firm-to-firm shipments within U.S., that on short distances the extensive margin explains almost the totality of the decay.

Again, however, we find significant differences across sectors. For Personal Services or Motion Pictures, for example, almost the totality of the fall is due to fewer people traveling outside; For Food Stores, on the other hand, fewer people traveling outside only explains half of the fall in expenditure occurring out-of-home. What accounts for the rest of the drop?

Figure 2 indicates that across all sectors, 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. The drop in the average transaction value (the batch size margin) has a limited role in most cases. Tables B.7 and B.8 in the Appendix (p. 46) show that the combination of the account and frequency margin typically contribute 90-95% of the decline in expenditure.

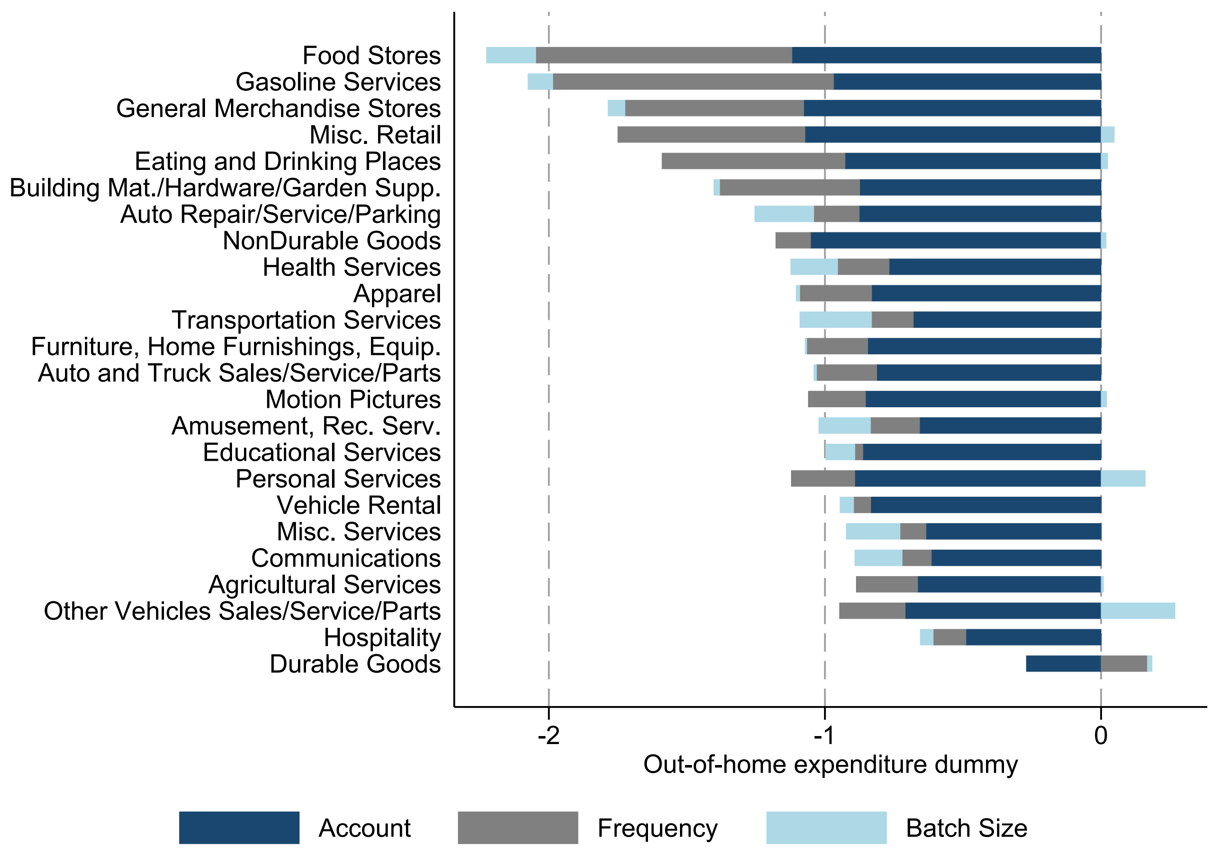

Figure 2: Margins in the out-of-home expenditure drop

Figure 2 also suggests that a large part of the heterogeneity in gravity seems associated to heterogeneity in the frequency margin. 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.[13] 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.

Figure 3: Drop in expenditure out of home

Our stylized model below will provide a possible explanation for this correlation, based on heterogeneity in storage costs across sectors. When storage costs are high, consumers want to reduce the average inventory held. To do so, they need to purchase smaller batches more frequently. Since travel is expensive, however, they reduce the average distance traveled. 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 matter in determining local equilibrium outcomes. In sectors where storage costs are high, consumer demand declines faster with distance: hence, from a merchant's point of view, the market will be more

[13]Figures B.3, B.4 and B.5 in the Appendix, starting at page 48, replicates it for all the sectors and for the impact of distance using gravity.

localized, and distant competition will be less of a threat. The evidence, however, is purely cross–sectional, and the observed correlation might be spuriously induced by unobserved supply–side characteristics of the sector. In the next section we consider a more direct shock to travel cost.

3 The Effect of Rain

The analysis up to now has shown that consumers' typical travel ranges are limited, expenditure declines with distance, and the combination of accounts and frequency margins are the main reasons of such decline. We have also shown that the strength of gravity varies by sector, and in a way that can be related to sector-level characteristics like storage costs and durability.

While certainly suggestive, these facts per se do not show that the consumer is actively managing the spatial dimension of consumption. Some sector-level characteristics of 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: everything else equal, high fixed costs would imply less dense suppliers, so that consumers would need to travel farther on average and expenditure would decline less with distance; since travel is expensive, however, a higher average distance will induce less frequent transactions, thus replicating the correlation in Figure 3.

To make progress on this issue, we need a plausible shifter to consumers' travel costs whose variation is uncorrelated to residential decisions of consumers and locations decisions of firms in the sample period. We can then study the impact of this cost shifter on the spatial distribution of observed expenditure.

We turn to rain. We use 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 or a shopping location in our data, we find the closest weather station among the roughly twelve thousand disseminated over the U.S. territory. In the transaction data, the median distance between a weather station and a merchant is 6.5 km (mean 7.3 km) and the median distance to a residence is 7.3 km (mean 8 km).

We use 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 stations recorded rain both in the residence and in the shopping location. During the sample period, 28% 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–location pairs over time. This is not the case. A regression of the weather status indicator variable on residence – shopping location pairs and transaction date fixed effects absorbs only 24% of the variation in the transaction level data, leaving ample residual variation to identify movements in the spatial distribution of expenditure.

We recompute for each pair of locations h,s two expenditure flows: one observed during non-rainy days, and one observed during rainy days. At this point, we are in a position to extend our analysis from equations (1) and (2) above to include the effect of rain. Specifically, we start by estimating:

$$\log X_{hs} = \alpha + \gamma^{(h)} + \gamma^{(s)} + \eta \times \mathbf{1}_{(h \neq s)} + \rho \times \mathbf{1}_{RAIN} + \mu \times \left(\mathbf{1}_{(h \neq s)} \times \mathbf{1}_{RAIN}\right) + \varepsilon_{hs} \qquad (5)$$

where $\mathbf{1}_{RAIN}$ is an indicator variable assuming the value of 1 if the observation refers to rainy days and zero otherwise. The presence of origin and destination fixed effects ensures that average levels of rain by location are not contributing to the identification.[14] Table 5 shows the estimated values of ρ (column 1), η (column 3) and μ (column 5).

Table 5: **Expenditure out of home place and rain**

	Rain		Out of home		Out of home × Rain		Obs.
Category	coeff	pv	coeff	pv	coeff	pv	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Food Stores	-0.60	0.00	-2.02	0.00	0.29	0.00	35,942
Gasoline Services	-0.54	0.00	-1.84	0.00	0.29	0.00	55,990
General Merchandise Stores	-0.60	0.00	-1.57	0.00	0.29	0.00	39,601
Misc. Retail	-0.55	0.00	-1.51	0.00	0.26	0.00	48,146
Eating and Drinking Places	-0.47	0.00	-1.36	0.00	0.20	0.00	46,545
Building Mat./Hardware/Garden Supp.	-0.65	0.00	-1.20	0.00	0.25	0.00	22,298
Auto Repair/Service/Parking	-0.50	0.00	-1.12	0.00	0.27	0.00	7,296
NonDurable Goods	-0.38	0.00	-0.86	0.00	0.09	0.45	2,277
Health Services	-0.51	0.00	-0.99	0.00	0.17	0.02	8,412
Apparel	-0.44	0.00	-0.94	0.00	0.20	0.00	21,857
Transportation Services	-0.14	0.43	-0.75	0.01	-0.12	0.52	1,493
Furniture, Home Furnishings, Equip.	-0.52	0.00	-0.90	0.00	0.19	0.01	17,445
Auto and Truck Sales/Service/Parts	-0.49	0.00	-0.83	0.00	0.25	0.00	11,519
Motion Pictures	-0.43	0.00	-0.90	0.00	0.19	0.00	3,861
Amusement, Rec. Serv.	-0.35	0.00	-0.88	0.00	0.11	0.26	4,470
Educational Services	-0.47	0.00	-0.90	0.00	0.22	0.22	1,564
Personal Services	-0.42	0.00	-0.76	0.00	0.12	0.02	8,880
Vehicle Rental	-0.44	0.00	-0.83	0.00	0.36	0.02	1,310
Misc. Services	-0.41	0.01	-0.91	0.01	0.16	0.39	1,044
Communications	-0.12	0.40	-0.57	0.00	-0.11	0.51	1,005
Agricultural Services	-0.44	0.00	-0.75	0.00	0.13	0.14	2,517
Other Vehicles Sales/Service/Parts	-0.58	0.00	-0.60	0.29	0.44	0.03	1,082
Hospitality	-0.46	0.01	-0.53	0.01	0.24	0.21	2,219
Durable Goods	-0.52	0.01	-0.14	0.72	0.15	0.53	592

Rain affects expenditure at home significantly for most sectors. In the median sector, expenditure at home on a rainy day is about $100 \cdot \exp(-0.47) = 63\%$ of the expenditure on a non–rainy day. Food Stores, Building Materials and Garden Supplies, and General Merchandise are the most impacted; Communications, Transportation Services are the least impacted. Note that the ratio of ρ between a sector at the 10% of impact and one at the 90% is around 1.7 (and the response of expenditure between the least and the most impacted sector varies by a factor of 5): a common cost shock to all sectors induces differential responses across sectors.

[14] In this residence location – shopping location – weather status dataset, a regression of the weather status indicator on residence – shopping location pair fixed effects has an R^2 of 20%. The effect of rain is then identified comparing the same pair of locations in rainy and non–rainy days.

Rain also implies a drop in expenditure out of home: $\rho + \mu$ is negative for all sectors. In the typical sector, expenditure outside on rainy days is 77% of the expenditure outside on non–rainy days. The heterogeneity in the responses stays in the same order of magnitude: the $p90/p10$ ratio in $\rho + \mu$ is around 1.6, and the max/min ratio is around 4.5.

Not only does a cost shock impact the levels of consumers expenditure, but it does so differentially over space. Column 5 reveals that rain impacts the spatial distribution of expenditure making the decline in expenditure out of home *less* pronounced. This behavior is consistent with the shock selecting the type of agents going out for shopping. While some agents do not shop (the expenditure declines both at home and outside), those who choose to go out must be those less sensitive to travel costs, and hence the composition of expenditure shifts toward out-of-home locations. For the median sector, the out-of-home expenditure decline in rainy conditions is 19% flatter than in non-rainy days.

We also find that the spatial distribution of expenditure moves differentially across sectors. A simple statistic to consider would compare $\exp(\eta)$, the percentage drop in out-of-home expenditure vs. home expenditure without rain, to $\exp(\eta + \mu)$, giving the same percentage in rainy days. In Figure 4 we plot $\exp(\eta + \mu) - \exp(\eta)$ against the frequency of transactions.[15]

A common travel cost shock makes gravity flatter by only 4.4 percentage points in Food Stores; this flattening grows to 8.7 points in Motion Pictures, 14 points in Durable Goods and 18.6 points in Vehicle rentals. The spatial distribution of demand is impacted less for sectors with more frequent purchases. This finding provides some validation to Figure 3 above. When a good is less storable, purchases are more frequent but also more local; moreover, expenditure is less sensitive to rain episodes.

A final insight into the impact of rain can be gauged by comparing Table 5 with the results of the following gravity regression:

$$\log X_{hs} = \alpha + \gamma^{(h)} + \gamma^{(s)} + \delta \log dist_{hs} + \tilde{\rho} \times \mathbf{1}_{RAIN} + \tilde{\mu} \times (\mathbf{1}_{RAIN} \times \log dist_{hs}) + \varepsilon_{hs} \tag{6}$$

This regression augments eq. (2) with the weather status indicator, in levels and interacting it with distance. It estimates the effect of rain by comparing only places out of home at different distances. Table 6 reports the results. The most notable difference with respect to Table 5 above is that the interaction between distance and rain is now very small and almost everywhere statistically insignificantly different from zero. These findings suggest that rain alters the spatial composition of expenditure, but only by leaving at home those whom would take short trips *more* than those willing to travel longer: once expenditure is no longer occurring at very short distances, an extra kilometer of travel does not matter.

In this section, we are identifying a reasonable shock to individual travel costs: rain is actually affecting expenditure levels (ρ and μ are different from zero). Our finding are consistent with consumers actively

[15] This figure excludes one outlier, Other Vehicle Sales, Services and Parts, which is imprecisely estimated. This sector includes items like boats, motorcycles, or camper dealers, for example. For this sector, the estimated difference is 30 percentage points. Figure B.6 in Appendix, p. 49, shows the complete picture.

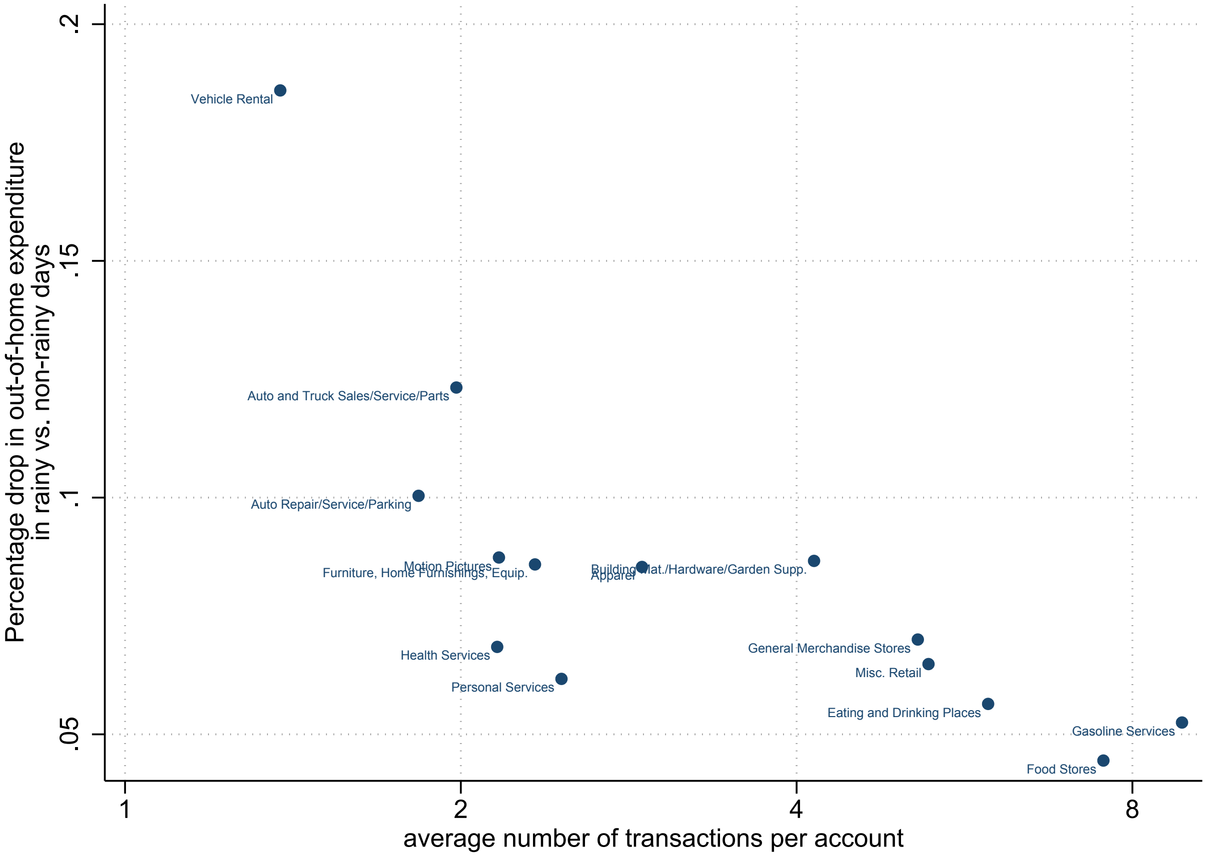

Figure 4: The flattening in gravity across sectors

managing the spatial dimension of their demand: if rain impacted expenditure at home and outside in the same proportion ($\rho \neq 0$, $\mu = 0$), this would mean that distance itself doesn't play a role in consumers' expenditure patterns; we find however that space does matter. Finally, we find that these consumer responses generates a pattern of responses systematically related to the frequency of transactions. If space mattered to the same degree across all sectors (identical μ for all the sectors), this would be an indication that the heterogeneity in distance travelled is mostly driven by heterogeneity in supply considerations; however, when we keep the supply fixed and compare origin–destination pairs over short periods of time, the response to an identical travel cost shock is smaller in sectors which are more frequently purchased. If those are also the sectors with high storage costs, this suggests that demand changes less in sectors where consumers hold on average lower inventories.

While this exercise points to consumers actively choosing how far and how frequently to travel, it is not able to identify the causal effect of changes in travel costs on overall expenditure: hence, we cannot use it to understand whether optimal consumer behavior has an overall impact on local equilibrium outcomes. We turn to this question next.

Table 6: **Gravity and rain**

Category	Rain		Distance		Distance × Rain		Obs.
	coeff	pv	coeff	pv	coeff	pv	
Food Stores	-0.40	0.00	-0.71	0.00	-0.00	0.98	28,485
Gasoline Services	-0.43	0.00	-0.50	0.00	0.03	0.00	47,006
General Merchandise Stores	-0.59	0.00	-0.77	0.00	0.07	0.00	34,539
Misc. Retail	-0.39	0.00	-0.50	0.00	0.01	0.45	41,449
Eating and Drinking Places	-0.44	0.00	-0.45	0.00	0.04	0.00	40,855
Building Mat./Hardware/Garden Supp.	-0.48	0.00	-0.56	0.00	0.01	0.69	17,873
Auto Repair/Service/Parking	-0.00	0.99	-0.28	0.00	-0.09	0.07	4,733
NonDurable Goods	-0.23	0.25	-0.48	0.00	-0.03	0.67	1,659
Health Services	-0.19	0.13	-0.18	0.00	-0.06	0.20	6,234
Apparel	-0.37	0.00	-0.40	0.00	0.03	0.13	18,979
Transportation Services	-0.23	0.35	-0.34	0.00	-0.02	0.85	1,251
Furniture, Home Furnishings, Equip.	-0.49	0.00	-0.41	0.00	0.05	0.10	14,964
Auto and Truck Sales/Service/Parts	-0.30	0.02	-0.20	0.00	0.01	0.80	8,290
Motion Pictures	-0.16	0.22	-0.21	0.00	-0.05	0.36	2,393
Amusement, Rec. Serv.	-0.38	0.01	-0.18	0.00	0.04	0.40	3,399
Educational Services	-0.44	0.15	-0.15	0.31	0.06	0.54	1,106
Personal Services	-0.21	0.02	-0.20	0.00	-0.04	0.25	6,299
Vehicle Rental	-0.01	0.98	-0.03	0.83	-0.04	0.70	741
Misc. Services	0.11	0.75	1.16	0.00	-0.15	0.19	602
Communications	0.35	0.28	-0.27	0.04	-0.22	0.04	634
Agricultural Services	-0.47	0.02	0.64	0.00	0.06	0.45	1,350
Other Vehicles Sales/Service/Parts	-0.15	0.69	-0.45	0.09	-0.01	0.97	702
Hospitality	-0.17	0.55	-0.11	0.12	-0.02	0.82	1,855
Durable Goods	-0.35	0.40	1.70	0.30	-0.01	0.94	365

4 Consumers Demand and Local Outcomes

Are consumers' spatial consumption patterns important to equilibrium local economic activity? One way to answer this question is to study the impact of (plausibly exogenous) differences in local population size on local sectoral employment, as a function of demand-related sector characteristics. Our analysis has shown that gravity drops faster in sectors which are more frequently purchased; moreover, a common rain shock impacts those sectors less. We have argued that differences in the average frequency of transactions across sectors may reflect differences in the storability of products. One would expect that everything else equal, higher storage costs make demand more "local": consumers would rather buy smaller batches and in more frequent trips; since travel is expensive, however, consumers should optimally choose to take shorter trips. Hence, in response to a larger local population, we should expect to see local employment in high storage cost sectors to be larger than employment in sectors with low storage costs. Moreover, if savings in travel time are at the root of this behavior, we should expect to see employment growth to be driven by a higher density of stores (that is, a reduced average distance between consumers and stores), rather than more employees per store.

In the next subsection, we study the impact of differences in population on local sectoral employment as a function of the sector's average number of transactions, which we think of as a simple proxy for storage costs. The advantage of this approach is that the proxy is not filtered through functional form assumptions. On the other hand, this statistic may be contaminated by sectoral characteristics other than storability. Since direct measures of storability are not available, we then propose a highly stylized model of consumer behavior to recover a parameter that measures the storage costs of products. The purpose of that model is not to construct an equilibrium theory of consumer purchase, but to address some of the unobserved factors that may influence the observed frequency of transactions. In both cases, we show that in response to a larger population, employment is larger in high storage costs sectors, and that this growth is driven by a higher density of stores.

4.1 Employment and frequency of transactions

We start by proxying sectoral storage costs with the average frequency of transactions in each sector in the sample. To explore the heterogeneous response of local sectoral outcomes to local population we estimate,

$$\ln y_{sct} = \alpha + \beta \ln pop_{ct} + \gamma \ln freq_s \times \ln pop_{ct} + \eta_0 \ln i_{ct} + \eta_1 \ln size_c + FE + \varepsilon_{sct} \tag{7}$$

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 from 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 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 studies total local employment as an outcome. Column (1) shows the OLS estimate of eq. (7) in the cross section of counties in 2007. As expected, larger income and population enter positively. However, sectors with higher storage costs (as proxied with a higher frequency of transactions) have a smaller employment in the cross section of counties. Obviously, population and sectoral employment may be correlated via a number of unobserved factors, which renders the coefficient on the interaction between population and frequency hard to interpret. To circumvent such omitted variable problem, we instrument county population with information on the underlying geological composition of the county. In particular, we follow ideas developed in Burchfield, Overman, Puga, and Turner (2006) and Duranton and Turner (2017) and instrument population with the county composition of different types of aquifers on which the territory of a county lays.[16] We also instrument the interaction of population and frequency with the

[16] Burchfield, Overman, Puga, and Turner (2006) and Duranton and Turner (2017) use instruments based on this data as

interaction of the same percentage indicators with frequency. This strategy results in a robust first stage across all specifications. It is worth emphasizing that we see the forces we describe playing out in a long run equilibrium, and importantly, after the entry–exit margin of new establishments has been allowed to adjust. Our instrument, which moves county population in a cross sectional dimension, is consistent with this approach.

Columns (2)-(6) reports the results using this instrumental variable strategy. Column (2) shows that, after controlling for endogeneity, the sign on the interaction coefficient reverses. In counties where the population is allowed to be larger for underlying geological reasons, sectors with high storage costs have larger employment in 2007 than sectors with low storage costs. Moving from the minimum to the maximum average frequency changes the growth in employment by 0.12 points, or 13.4% of the baseline impact of population. Column (3) shows the same regression run in 1998: the effect of storage costs on net employment is positive but insignificant: we will show below that this is the result of two forces pushing strongly in opposite directions. To assess the stability of our estimates, we stack our two cross-sections and add year fixed effects in column (4): the coefficient turns back to significance. Obviously, time trend may be operating differentially for different states and sectors, and this may affect our estimates in the stacked regression. In column (5) and (6), we allow for heterogeneous time trends across sectors (both columns), and across U.S. States (column (5)) or commuting zones (column (6)). Our estimated coefficient becomes a little smaller, but stays significant. In the most restrictive specification, moving from the smallest to the largest frequency changes the employment response by 0.07, or 5.2% of the baseline impact of population.

The regressions shown so far indicate that if storage costs are well proxied by the observed frequency of transactions, consumers spatial optimization of their purchases has relevant consequences on economic outcomes. To provide evidence that this equilibrium impact is realized by our mechanism, we ask next how is employment increasing in the county. In principle, an increase in local sectoral employment may be generated entirely at the intensive margin, i.e., via more employees per store. If storage costs are important, however, we should expect that demand in high storage costs sectors is served by more geographically concentrated stores, i.e., via a lower average distance between consumers and stores. Table 8 and 9 show that this is indeed the case. In particular, Table 8 replicates Table 7 but uses the log number of establishments in a given county-sector-time as a dependent variable. Estimates on the interaction coefficient are now always very large and strongly significant, and the coefficient is fairly stable. This table shows that in sectors with higher storage costs, the number of stores grows faster: this is 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 increasing employment via a denser presence of stores. Table 9 shows that, if anything, stores become relatively smaller on average: this is again consistent with a situation where their is a larger demand for a denser supply in higher storage costs sectors. In particular, the lack of employment response in 1998 is due to a relative increase in the number

an exogenous shifter for density of population. We find that the percentage composition of a county area laying on top of diffent types of aquifers results in a good first stage to exogenously shift population.

Table 7: **Local employment and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	county-sector log employment					
Sample years :	07	07	98	98,07	98,07	98,07
Log land area	0.096***	0.090***	0.110***	0.099***	0.130***	-0.020
	(0.005)	(0.005)	(0.005)	(0.004)	(0.008)	(0.016)
Log income per capita	0.994***	1.285***	1.357***	1.341***	1.342***	1.007***
	(0.020)	(0.041)	(0.050)	(0.032)	(0.065)	(0.076)
Log population	1.208***	0.943***	1.033***	0.976***	1.000***	1.252***
	(0.007)	(0.040)	(0.040)	(0.029)	(0.035)	(0.035)
Log population × log frequency	-0.060***	0.067**	0.018	0.046**	0.040**	0.035*
	(0.005)	(0.028)	(0.029)	(0.020)	(0.020)	(0.019)
Sector Fixed Effects	Yes	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	No	Yes
R-square	0.84	0.83	0.84	0.84	0.84	0.85
N	49,876	49,876	50,263	100,139	100,139	100,139
F-stat: Log population		71.3	91.1	152.3	66.8	157.1
F-stat: Log population × Log frequency		56.6	53.5	109.5	114.2	138.9

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

of stores that is almost completely offset by a relative decrease in the number of employees per store.

4.2 A Simple Model of Shopping

In this section we develop a highly stylized model of consumer behavior. The model will only study the cross-sectional implications of cost-minimizing consumers. Our purpose is simply to extract from the data a measure of storage costs that remove some of the main concerns about the simple use of frequency of transactions. We generalize ideas present in Oi (1992) to a setting where consumers with heterogeneous travel costs choose how far to travel for their purchases.

There is a continuum of locations index by $j \in [0, +\infty)$. Consumers are indexed by ω and live in location 0. An agent ω wants to buy a fixed amount $\bar{q}(\omega)$ and consumes it linearly in one unit of time. There is one merchant in any location j. The price schedule is fixed at $p(j)$, with $p'(j) < 0$: farther places have lower unit prices. However, longer travels are more expensive, and storage is costly: consumers trade off savings from distance with extra travel and storage costs.

If consumers choose to buy a batch of size z per trip, they will have to make $\bar{q}/z$ trips overall (integer constraints are ignored). Consumers choose how far to travel (j) and how much to buy per trip (z), given their individual travel cost $t(\omega)$, sector–specific storage costs g, and a travel cost function; in particular,

Table 8: **Number of establishments and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	county-sector log number of establishments					
Sample years :	07	07	98	98,07	98,07	98,07
Log land area	0.102***	0.095***	0.122***	0.108***	0.127***	0.034***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.010)
Log income per capita	0.877***	1.215***	1.258***	1.259***	1.162***	0.914***
	(0.014)	(0.028)	(0.033)	(0.022)	(0.042)	(0.047)
Log population	0.871***	0.566***	0.608***	0.574***	0.625***	0.786***
	(0.004)	(0.025)	(0.025)	(0.018)	(0.022)	(0.022)
Log population × log frequency	0.009***	0.154***	0.137***	0.149***	0.134***	0.131***
	(0.003)	(0.018)	(0.018)	(0.013)	(0.013)	(0.012)
Sector Fixed Effects	Yes	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	No	Yes
R-square	0.89	0.87	0.88	0.87	0.88	0.90
N	49,876	49,876	50,263	100,139	100,139	100,139
F-stat: Log population		71.3	91.1	152.3	66.8	157.1
F-stat: Log population × Log frequency		56.6	53.5	109.5	114.2	138.9

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

Table 9: **Number of employees per establishment and frequency of purchase**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	county-sector log number of employees per establishment					
Sample years :	07	07	98	98,07	98,07	98,07
Log land area	-0.002	-0.005	-0.012***	-0.009***	0.003	-0.054***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.006)	(0.012)
Log income per capita	0.164***	0.070**	0.100***	0.082***	0.179***	0.093
	(0.013)	(0.029)	(0.037)	(0.023)	(0.048)	(0.057)
Log population	0.327***	0.377***	0.424***	0.402***	0.375***	0.466***
	(0.004)	(0.029)	(0.030)	(0.021)	(0.026)	(0.027)
Log population × log frequency	-0.065***	-0.087***	-0.119***	-0.103***	-0.093***	-0.097***
	(0.004)	(0.021)	(0.022)	(0.015)	(0.015)	(0.015)
Sector Fixed Effects	Yes	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	No	Yes
R-square	0.48	0.53	0.55	0.54	0.55	0.56
N	64,244	49,876	50,263	100,139	100,139	100,139
F-stat: Log population		71.3	91.1	152.3	66.8	157.1
F-stat: Log population × log frequency		56.6	53.5	109.5	114.2	138.9

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

the consumer solves,

$$c(\omega) = \min_{j,z} p(j)\,\bar{q}(\omega) + \kappa(j; t(\omega))\,\frac{\bar{q}(\omega)}{z} + g\frac{z}{2} \tag{8}$$

where $\kappa(j, t(\omega))$ describes the cost of traveling a distance j for an individual with travel cost $t(\omega)$. In this expression, $z/2$ is the average inventory held given that the consumer depletes its inventory at a linear rate. We assume that $\ln t(\omega)$ and $\ln \bar{q}(\omega)$ are uncorrelated within consumers and have mean zero across consumers.[17] Note that

$$X = p(j)\,\bar{q}(\omega) \tag{9}$$

is the account–level expenditure, which we measure in the credit card data.

The optimal batch size is such that,

$$z(j; t(\omega)) = \left(\frac{2\bar{q}\kappa(j; t(\omega))}{g}\right)^{1/2} \tag{10}$$

and hence a consumer travelling to j will optimally choose a number of trips equal to

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

Conditional on travelling a distance j, when travel costs are larger consumers will take less frequent trips and buy a larger batch per trip. Also, when storage costs are higher consumers will take more frequent trips and buy smaller batches per trip.

Substituting these expressions in the cost function,

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

A marginally longer distance travelled makes consumers save $p'(j)$ per unit purchased; however, consumers pay more in travel costs and sustain larger storage costs (since they optimally buy larger batches). To gain intuition, we assume $\kappa(j; t(\omega)) = t(\omega)\,j^{\alpha}$, and $p(j) = j^{-\beta}$. The first order condition with respect to j requires,

$$j = \left[\left(\frac{\beta}{\alpha}\right)^2 \frac{2\bar{q}(\omega)}{g}\right]^{1/(\alpha+2\beta)} t(\omega)^{-1/(\alpha+2\beta)} \tag{13}$$

A consumer will optimally choose to travel a shorter distance if his idiosyncratic travel costs are larger, of if the storage costs are higher.

[17] This simple model is closely related to monocentric city models. In a monocentric city model, workers on a line commute to a single Central Business District: they choose where to live trading off lower rents with higher commuting costs. Similarly, here consumers live concentrated in a particular place and choose how far to travel trading off lower prices with higher travel and storage costs.

4.3 Empirical Implications

Note that the frequency equation (11) implies

$$\ln f_s(\omega) = \ln\left(\frac{g_s}{2}\right)^{1/2} - \frac{\alpha_s}{2}\ln j(\omega) - \frac{1}{2}\ln\frac{t(\omega)}{\bar{q}_s(\omega)} + \ln\tilde{\varepsilon}_s(\omega) \tag{14}$$

We have added sector s indices to indicate that storage costs and estimated travel costs may vary across sectors and that quantities purchased may vary within individuals across sectors. This equation says that conditional on a person's characteristics, the elasticity of the frequency of trips with respect to distance should identify α_s. We estimate this equation sector by sector. To estimate it, we start from the transaction–level dataset and compute 20 quantiles of distance travelled δ, and the average distance within the quantile $\bar{j}_\delta$. For each account, we then compute the frequency of trips taken at a given distance quantile, $f_{\delta s}(\omega)$, and estimate

$$\ln f_{\delta s}(\omega) = \gamma_0 - \frac{\alpha_s}{2}\ln\bar{j}_{\delta s} + I_s(\omega) + \varepsilon_{\delta s}(\omega) \tag{15}$$

where $I_s(\omega)$ are account fixed effects. This specification allows for unobserved heterogeneity in consumers characteristics and hence controls for a variety of factors like income, location of residence and correlated characteristics, family size, transportation availability; hence, it removes omitted variable bias arising from consumers' characteristics endogenously determining distance. Since this equation is estimated by sector, the estimated shape of travel costs will also capture differences in supply density across sectors: in denser sectors, consumers need to travel less to access a given number of suppliers, and hence the regression will measure a more negative α_s.

From the observed expenditure equation (9),

$$\ln X_s(\omega) = -\beta_s \ln j(\omega) - \ln\bar{q}_s(\omega) + \ln\tilde{\varepsilon}_s(\omega)$$

This equation says that, conditional on individual characteristics, the elasticity of expenditure to distance identifies β_s. This statement relies on the fixed quantity assumption; in the estimated equation, β_s will also capture differences in the elasticity of demand and hence the degree of differentiation of products within the sector. As above, we estimate this equation sector by sector, and run

$$\ln X_{\delta s}(\omega) = \gamma_1 - \beta_s \ln\bar{j}_{\delta s} + I_\delta(\omega) + \ln\varepsilon_s(\omega)$$

where $X_\delta(\omega)$ is the expenditure of account ω on locations in the distance bin δ, $\bar{j}_\delta$ is the average distance in the bin, $I_\delta(\omega)$ is a person fixed effect and $\varepsilon_s(\omega)$ is a stochastic error term. Again, the presence of account fixed effects allows for unobserved individual characteristics and controls for endogeneity.

Consider finally the distance equation (13):

$$\ln j_s(\omega) = \frac{1}{\alpha_s + 2\beta_s}\ln\left[\left(\frac{\beta_s}{\alpha_s}\right)^2 \frac{2}{g_s}\right] + \frac{1}{\alpha_s + 2\beta_s}\ln\frac{\bar{q}_s(\omega)}{t(\omega)} + \tilde{\varepsilon}_s(\omega) \tag{16}$$

Note that the second and third term have an expected value of zero. This equation says that if all the sectors had homogeneous characteristics in terms of α_s and β_s (e.g., identical density of suppliers and elasticity of demand), differences in the average distance travelled by sector would identify differences in storage costs: a higher storage cost implies a lower average distance travelled across all accounts, everything else equal. This is in general not the case, as α_s and β_s vary by sector. We then estimate, pooling across all accounts and sectors,

$$\ln \bar{j}_s(\omega) = \sum_{\sigma=1}^{S} 1_{\sigma=s} \cdot \tau_\sigma + \varepsilon_s(\omega) \tag{17}$$

where $\bar{j}_s(\omega)$ is the average distance travelled by the account in sector s, $1_{\sigma=s}$ is a sector fixed effect, and $\varepsilon_s(\omega)$ is a mean zero error term. The coefficients on the fixed effects in this regression identify

$$\tau_s = \frac{1}{\alpha_s + 2\beta_s} \ln\left[\left(\frac{\beta_s}{\alpha_s}\right)^2 \frac{2}{g_s}\right] \tag{18}$$

from which we can recover

$$\ln g_s = \ln\left[2\left(\frac{\beta_s}{\alpha_s}\right)^2\right] - (\alpha_s + 2\beta_s)\,\tau_s \tag{19}$$

In Figure 5, we plot our estimates of storage costs against the log frequency of transactions in the sector. The figure broadly confirms our intuition, that sectors with higher frequency of transactions tend to have higher storage costs. Also, as one would expect, some services tend to be measured as high storage costs sectors even if they have lower frequency of transactions. In the section that follows, we use this storage cost to study the impact of local population differences on local sectoral employment and store density.

4.4 The Impact of Storage Costs on Local Equilibrium Outcomes

In this section, we test whether the intuition emerging from the results in Section 4.1 carries through with a model–generated measure of storage costs. We now estimate

$$\ln y_{sct} = \alpha + \beta \ln pop_{ct} + \gamma \ln g_s \times \ln pop_{ct} + \eta_0 \ln i_{ct} + \eta_1 \ln size_c + FE + \varepsilon_{s,c,t} \tag{20}$$

Table 10 shows the response of local sectoral employment to exogenous differences in population as a function of storage costs. The pattern of significance is broadly similar and it improves slightly in column (6). In the most restrictive specification of column (6), moving from the smallest to the largest storage cost increases the employment impact by 0.115 points, or about 9% of the baseline impact of population. Table 11 shows again that the increase in employment comes about via a (larger) increase in the density of establishments. This makes sense within the context of the stylized model above: when storage costs are higher, consumers want to take more frequent trips, but more close-by; hence, a larger population should imply a demand that is more geographically concentrated in high storage cost sectors than in low storage

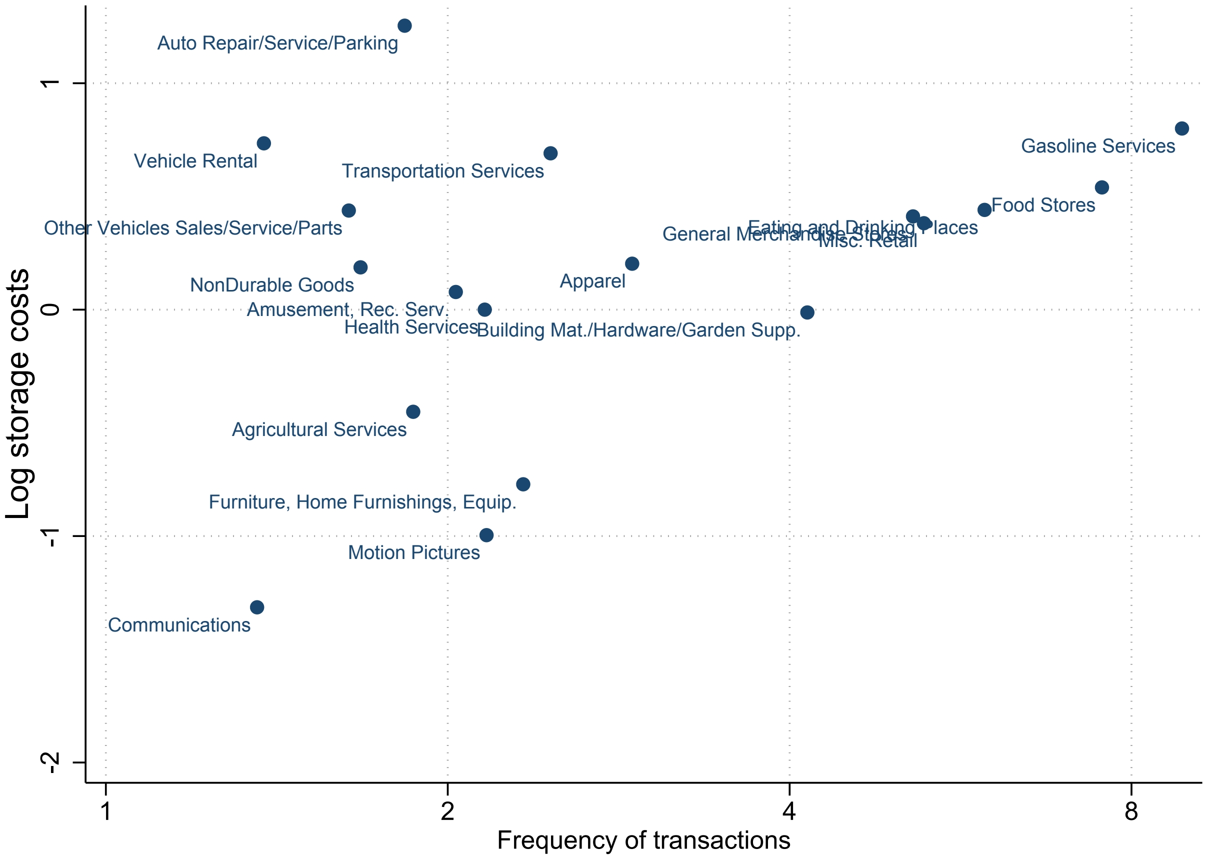

Figure 5: Estimated storage costs against frequency of transactions (log scale)

costs sectors. In the most restrictive specification, a 1% increase in population increases the number of stores by 0.77% in the lowest storage cost sector, but by 1.05% in the highest storage cost one. As anticipated by our regressions above, we confirm that this increase in the number of stores comes partially at the expenses of store size, which grows more slowly: column (6) in table 12 shows that a 1% increase in population increases employment per establishment by 0.45% in the low storage costs sectors, but only by 0.29% in the highest storage cost one. Again, the regressions show that the lack of employment response in 1998 is due to the effect on the number of stores roughly compensating the effect on the number of employees per store.

Taken together, these results paint a consistent picture of the importance of spatial consumption patterns for local economic outcomes. In sectors with high storage costs, consumers are more willing to trade off larger batches with frequent trips, but to do so they choose to travel shorter distances. We start by taking the frequency of trips as a proxy for storage costs, and have shown that in the raw data it is the case that exogenous differences in population across U.S. counties generate increases in local employments which are larger in higher storage costs sectors. The advantage of using such a simple proxy is that it is untouched by modelling assumptions. The disadvantage, as a simple model shows, is that the average

Table 10: **Local employment and storage costs**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	county-sector log employment					
Sample years :	07	07	98	98,07	98,07	98,07
Log land area	0.096***	0.090***	0.110***	0.099***	0.131***	-0.021
	(0.005)	(0.005)	(0.005)	(0.004)	(0.008)	(0.016)
Log income per capita	0.994***	1.282***	1.362***	1.340***	1.343***	1.003***
	(0.020)	(0.040)	(0.049)	(0.032)	(0.066)	(0.076)
Log population	1.152***	1.008***	1.047***	1.019***	1.036***	1.285***
	(0.003)	(0.017)	(0.016)	(0.012)	(0.022)	(0.027)
Log population $\times$ log storage costs	-0.059***	0.065**	0.031	0.049**	0.048**	0.045**
	(0.005)	(0.031)	(0.030)	(0.022)	(0.022)	(0.021)
Sector Fixed Effects	Yes	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	No	Yes
R-square	0.84	0.83	0.84	0.84	0.84	0.85
N	49,876	49,876	50,263	100,139	100,139	100,139
F-stat: Log population		400.1	482.8	837.4	264.3	238.3
F-stat: Log population $\times$ log storage costs		80.6	71.3	151.4	145.5	138.6

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

Table 11: **Number of establishments and storage costs**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	county-sector log number of establishments					
Sample years :	07	07	98	98,07	98,07	98,07
Log land area	0.102**	0.094***	0.121***	0.107***	0.127***	0.034***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.010)
Log income per capita	0.875**	1.204***	1.251***	1.250***	1.154***	0.905***
	(0.013)	(0.027)	(0.033)	(0.021)	(0.042)	(0.047)
Log population	0.883**	0.725***	0.739***	0.722***	0.757***	0.917***
	(0.002)	(0.011)	(0.011)	(0.008)	(0.014)	(0.017)
Log population $\times$ log storage costs	-0.017**	0.089***	0.138***	0.115***	0.109***	0.108***
	(0.003)	(0.019)	(0.020)	(0.014)	(0.013)	(0.012)
Sector Fixed Effects	Yes	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	No	Yes
R-square	0.89	0.87	0.87	0.87	0.88	0.90
N	49,876	49,876	50,263	100,139	100,139	100,139
F-stat: Log population		400.1	482.8	837.4	264.3	238.3
F-stat: Log population $\times$ log storage costs		80.6	71.3	151.4	145.5	138.6

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

Table 12: **Number of employees per establishment and storage costs**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	county-sector log number of employees per establishment					
Sample years :	07	07	98	98,07	98,07	98,07
Log land area	-0.006	-0.005	-0.012***	-0.008***	0.004	-0.055***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.006)	(0.012)
Log income per capita	0.119**	0.078***	0.111***	0.090***	0.188***	0.097*
	(0.015)	(0.029)	(0.036)	(0.023)	(0.048)	(0.057)
Log population	0.268**	0.284***	0.308***	0.297***	0.279***	0.368***
	(0.003)	(0.012)	(0.012)	(0.009)	(0.016)	(0.021)
Log population × log storage costs	-0.042**	-0.024	-0.107***	-0.065***	-0.061***	-0.064***
	(0.004)	(0.022)	(0.023)	(0.016)	(0.016)	(0.016)
Sector Fixed Effects	Yes	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	No	Yes
R-square	0.53	0.53	0.55	0.54	0.55	0.56
N	49,876	49,876	50,263	100,139	100,139	100,139
F-stat: Log population		400.1	482.8	837.4	264.3	238.3
F-stat: Log population × log storage costs		80.6	71.3	151.4	145.5	138.6

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

frequency may also be contaminated by factors like local density of suppliers or differences in demand elasticities. Hence, we turn to a simple model of cost–minimizing consumers, and recover storage costs by imposing some structure on the data. We show that even imposing this structure, storage costs continue to appear an important determinant of differences in local employment. Moreover, consistently with the model, these differences manifest themselves in more stores locally and slower growth in employment per store.

5 Discussion

In the last section we have shown that consumers' behavior arising from the purchasing technology contributes to shape local equilibrium outcomes. In this section we discuss some further implications and some possible shortcomings of our analysis.

Economic significance. Our estimates imply economically relevant adjustments. In the most conservative specification using model–based storage costs, a 10% increase in county population generates an increase in employment varying between 12.2% to 13.3%, as we move from the lowest to the highest storage cost in the sample. The heterogeneity in the impact is about 10% of the baseline coefficient on population. Associated to this variability, our estimates predict an increase in the number of stores varying from 7.7% to 10.5% (about 30% of baseline coefficient), and an increase in employment per store varying from 4.5% to 2.8% (or -44% of the baseline) as we move from low to high storage costs sectors.

Home market effects and theories of land use. The fact that sectoral employment increases faster than population is suggestive of home market effects mechanisms (e.g., Krugman, 1980). In our case, however, this effect is generated by the existence of a fixed cost of purchase that does not vary with volume, and heterogeneous storage costs playing the role of "transportation" cost in new economic geography model: when storage costs are high, demand is concentrated, and firms locate closer to consumers (i.e., they become denser, thereby reducing the average distance to their customers). Our results are also reminiscent of intuitions that can be traced back to Von Thunen (1826)'s model of rural land use.[18] In his theory, a central city is surrounded by (concentric circles of) rural land with possibly different uses: the transportation costs and perishability of different agricultural products to the central market will determine how far from the city different products will be produced.[19]

Fixed costs. In Tables 7-9 we have argued that an exogenous increase in population tends to generate a demand that is more geographically concentrated for high storage costs sectors than low storage costs ones. However fixed costs varying by industry may generate a denser presence of establishments in low fixed-costs industries: this heterogeneity could then replicate the negative relation between gravity and frequency. In that case, Tables 7-9 would be simply picking up the spurious correlation between frequency and gravity induced by underlying variation in fixed costs.

As a first response to this possible threat, we have proposed an estimate of the storage costs that attempts to account for these fixed costs. As an additional response, we can find proxies for sector-level differences in fixed costs directly, and introduce them as additional regressors in our empirical analysis. As for storage costs, direct observations of fixed costs are hard to obtain. However, a reasonable proxy is the national average of the ratio of total employees to total establishments in a sector–year: we will refer to this ratio simply as "fixed costs". If fixed costs are high, increasing returns to scale are more important and we should expect a higher employees to establishment ratio.

We preliminarily notice that at sector level, the correlation between log average frequency of transactions and log fixed costs is in fact equal to 0.12, and insignificantly different from zero: hence, it is not empirically true that low fixed cost sectors are those where transactions are more frequent.[20] We introduce the interaction of this measure with population and re-estimate the set of equations in (20). This new interaction variable is again instrumented with the interaction between fixed costs and county geological composition; moreover, the level of fixed costs is again absorbed by the sector-year fixed effects. If fixed costs are driving the results, we should expect that the interaction of population with log storage cost lose significance, and the interaction of fixed costs has a *negative* sign: an exogenous increase in population should increase county employment less in the sector which is sparsely (rather than densely) present, because consumers visit local stores more for the denser sectors (rather than the sparser one).

Columns (1) and (2) in Table 13 replicate the most conservative specifications in columns (5) and (6)

[18] See for example Von Thunen (1966).

[19] Interestingly, more perishable products like dairy should have been the ones produced closer to the market. For an empirical test of the Von Thunen model, see for example Fafchamps and Shilpi (2003) about Nepal.

[20] In our sample of 18 industries, we can compute fixed costs for 1998 and 2007. If we run a regression of log frequency on log fixed costs, estimated log storage costs and year dummies (36 observations), we find a coefficient of 0.17 (p.v. 0.29) for log fixed costs and a coefficient of 0.33 (p.v. 0.032) for log storage costs.

in Table 10 (the first four columns behave similarly and are omitted for brevity). The coefficient on the interaction between storage costs and population stays positive and in fact grows slightly in magnitude. Columns (3) and (4) replicates columns (5) and (6) of Table 11, where the dependent variable is the log number of establishments: controlling for fixed costs makes the density margin emerge even more strongly. The larger net effect on employment is also driven by a slower decline in employees per store, as reported in columns (5) and (6) of Table 13 (which compare to the last two columns of Table 12). The regression results show that, in response to differences in population, sectors with higher fixed costs also grow faster. However, as one would expect, this growth is for the major part driven by a higher number of employees per store (i.e., larger stores), than by more stores.[21] Overall, we read these results as further evidence that consumers' mobility impact local economic outcomes.

Table 13: **Local outcome responses and storage costs, 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	0.792***	1.042***	0.655***	0.817***	0.138***	0.225***
	(0.063)	(0.063)	(0.041)	(0.041)	(0.049)	(0.053)
Log population × log storage costs	0.066***	0.063***	0.117***	0.116***	-0.051***	-0.053***
	(0.023)	(0.021)	(0.014)	(0.013)	(0.017)	(0.017)
Log population × log fixed costs	0.092***	0.093***	0.038***	0.037***	0.055***	0.056***
	(0.022)	(0.021)	(0.014)	(0.014)	(0.019)	(0.018)
Log land area	0.131***	-0.021	0.128***	0.035***	0.003	-0.056***
	(0.008)	(0.016)	(0.006)	(0.010)	(0.006)	(0.012)
Log income per capita	1.348***	1.002***	1.164***	0.912***	0.184***	0.090
	(0.066)	(0.076)	(0.043)	(0.048)	(0.048)	(0.057)
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	0.84	0.85	0.88	0.89	0.55	0.57
N	100,139	100,139	100,139	100,139	100,139	100,139
F-stat: Log population	7.6	40.1	7.6	40.1	7.6	40.1
F-stat: Log population × log storage costs	152	143.8	152	143.8	152	143.8
F-stat: Log population × log fixed costs	26.7	36	26.7	36	26.7	36

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

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 of payments. 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 consumers' characteristics: hence, short trips are less likely to be reported in our data. On the one hand, this

[21] One might have expected that, following our argument above, sectors with high fixed costs grow less in terms of employment. Our objective in this paper is not to explain the heterogeneous response of employment with respect to fixed costs. We however speculate that the positive coefficient on the interaction between log fixed costs and population is driven by the fact that in the cross-section, fixed costs and frequency are not negatively but positively (and insignificantly) correlated.

selection will make gravity appear less important than it actually is, since we are removing expenditure falling close-by; this effect will be in fact stronger in sectors where the average distance traveled is shorter, i.e., in sectors with high frequency of transactions. Via this first channel, the relation between gravity and frequency documented in Figure 3 should be steeper than we can measure. On the other hand, this selection will also remove more of the short trips (which are higher frequency) than the longer trips (which are low frequency). Via this second channel, the relation should be flatter than we can measure. The fact that these two forces compensate each other is somewhat reassuring, but we cannot offer clear predictions on the net effect of these unobservable choices, and hence our results should be interpreted with this limitation in mind.

6 Conclusion

Using detailed geographical information from more than 1.7 million individual consumers' credit card transactions, we document several stylized facts regarding the geography of consumption. We find large heterogeneity across industries in the overall impact of distance and in the importance of extensive margins. We also find that the differences in gravity across industries are correlated with the frequency of transactions. A simple model of consumer choice suggests that this correlation can be induced by heterogeneity in the durability/storability of the final item demanded, which we interpret as a source of gravity. We argue that storable/durable final consumption items may have an inherently larger geographical markets than perishable/non-durable goods. Hence, the distribution of travel costs of consumers will matter differentially for firms operating in different sectors.

Since unobserved heterogeneity in sector characteristics might be a contributing factor in determining gravity patterns, we provide more direct evidence that consumers are actively managing the spatial dimension of their consumption using shocks to travel costs induced by rain. We find that rain reduces expenditure of consumers living in a particular place both at home and in outside locations, but differentially over space and across sectors. In particular, gravity becomes flatter, a behavior consistent with consumers most sensitive to travel costs to postpone their purchases; moreover, rains matters more for more durable/storable goods.

We finally show that consumers' mobility has implication for local 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.

Taken together, these results indicate that local demand conditions are quantitatively relevant for local outcomes, and contribute to determine their equilibrium response to local shocks and policies. Incorporating demand-side characteristics is important to analyzing firms' location and production decisions, in particular in the understudied service sector, which accounts a large share of economic activity. Further, our results provide important information for the study of the liberalization of international trade and investment in services: entry and location decisions of foreign firm establishments in a local market will be shaped, among other things, by the different degrees of localization of their product's market.

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, 100:5, pp. 2304-2339.

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

[5] Anderson, J. and E. Van Wincoop (2003). "Gravity with gravitas: a solution to the border puzzle", American Economic Review, 93:1, pp. 170-192

[6] Bernard, A., B. Jensen, S. Redding and P. Schott (2010). "Wholesalers and Retailers in U.S. Trade (Long Version)", NBER Working Paper 15660.

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

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

[9] Disdier, A. and K. Head (2008). "The puzzling persistence of the distance effect on bilateral trade", Review of Economics and Statistics, 90:1, pp. 37-48.

[10] Davis, D., J. Dingel, J. Monras, and E. Morales (2017). "How Segregated is Urban Consumption?", Working paper.

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

[12] 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, 67:5, pp. 823-839.

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

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

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

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

[17] Handbury, J., I. Rahkovsky, and M. Schnell (2016). "Is the Focus on Food Deserts Fruitless? Retail Access and Food Purchases Across the Socioeconomic Spectrum", Working paper.

[18] Hillberry, R. and D. Hummels (2007). "Trade responses to geographic frictions: A decomposition using micro-data", European Economic Review 52:3, pp. 527:550

[19] Hummels, D. and P. Klenow (2005). "The variety and quality of a nation's exports", American Economic Review, 95:3, pp. 704-723.

[20] Hummels, D. and A. Skiba (2004). "Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture", Journal of Political Economy, 112:6, pp. 1384-1402.

[21] 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, 1:1, p. 53-74.

[22] Hortaçsu, A. and C. Syverson (2015). "The ongoing evolution of US retail: A format tug-of-war", Journal of Economic Perspectives, 29:4, pp. 89-112.

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

[24] 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.

[25] Krugman, P. (1980). "Scale economies, Product Differentiation and the Pattern of Trade," American Economic Review, 70:5, pp. 950-959.

[26] 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, 29, pp. 897-910.

[27] Monte, F., S. Redding and E. Rossi-Hansberg (2016). "Commuting, migration, and local employment elasticities", Working paper.

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

[29] 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.

[30] Reilly, W. (1931). “The Law of Retail Gravitation”, New York: Knickerbocker Press.

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

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

[33] 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.

A Data Processing

A.1 Merchant codes

The transaction data classifies merchants using the MCC classification. Classification 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, 221 nested narrow ones.

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 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 observation, around 1.1 million records are related to 1) cash advances, interests, 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 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 2,207,907 transactions from 80,087 accounts.

A.3 Account data

The account data for the months of March to October 2003 has originally 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, updates to the month end balance, an income or residence change, 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, 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 observations for 239,369 unique accounts.

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 2,189,048 transactions matched (more than 99% of the transactions data file) from 79,209 unique accounts. Out of the totality of matched transactions, only 155,254 did not find the exact event month in the account information: 145,815 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 6,147 transactions do not meet this requirement: inspection shows that 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 86% of the 6 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 2,154,927 observations.

To identify the set of observations that it may be possible to 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 9,004 observations with an empty potential location name.

We then mark transactions of common online providers[22] and find the words "Online", "On Line", ".com", ".net" in 102,715 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 206,012 of them. After this processing, we are left with 1,931,815 transactions that may contain city names, 90% of those for which a state name could be found, for 73,959 unique accounts. Note that the largest contributor to the drop

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

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 contain, 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.[23]

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 1,475,545 out of the 1,931,815 we intend to match, 76% of our observations. Out of the 456,270 transaction with no match, 123,861 have names and states that match the MCD list. We assign "match quality" equal to 10 to those transactions matched at this first pass. We have 332,409 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

[23] An alternative choice 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.

recover 117,787 observations, bringing the number of matched transactions to 1,604,485.

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 21,080 observations, bringing the number of matched transactions to 1,738,273.

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 33,700 observations more (also assigned "match quality" equal to 10), bringing the total to 1,771,973 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 1,751,067 observations (90.6% of the transactions we intended to match) and 71,927 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 Additional Empirical Results

B.1 Summary Statistics by State

Table B.1 shows summary statistics on our main dataset by State of transaction.

B.2 Frequent users

Our data is unfortunately sparse enough not to allow a full analysis of individual consumers' behavior, since the median account uses the credit card around once per month, and around 96% of the accounts use the credit card less than once every two days. Here we focus on consumers with at least 15 transactions per month on average, and on transactions within 120 km from a consumer's residence. We term this as "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 2,198 accounts, conducting 496,654 transactions over the sample period. They reside in 1,729 locations and shop in 6,930 of them; there are a total of 26,436 origin-destination combinations over which we observe transactions.

Table B.2 shows summary statistics for such sample. Consumers in the median residence visit only 13 distinct sales locations overall during the sample period (15.3 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 here has 231 sales locations within 120 km (compared to 186 for the whole data). Hence, the median residence sees consumers shop in 5% of the available locations (the mean is 7%), very comparable to the values in the general data (4% and 6% respectively).

B.3 Percentiles of distances traveled

These Tables show summary statistics on the percentiles of distances traveled by consumers, by sector. Table B.3 refers to percentiles in the unweighted distribution. Table B.4 shows the same percentiles weighting each transaction with the correspondent purchase value.

B.4 Gravity over all distances

In Figure B.1, we estimate eq. (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. The blue line in Figure B.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. Different sectors are more or less represented at different distances (see also Tables B.3 and B.4), implying that the coefficient δ varies.

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

State	Median	Mean	St. Dev.	Sum	N
AK	33	74	150	137,224	1,866
AL	28	64	173	1,090,956	17,060
AR	29	64	159	553,669	8,711
AZ	29	72	234	1,883,945	26,157
CA	31	74	213	11,058,192	149,493
CO	26	62	181	1,764,997	28,603
CT	31	71	191	4,238,450	60,117
DC	28	71	165	332,700	4,703
DE	30	73	217	494,720	6,743
FL	30	73	216	7,588,112	104,520
GA	27	64	184	2,720,214	42,260
HI	35	85	221	463,444	5,426
IA	28	61	170	814,099	13,454
ID	29	67	165	317,258	4,747
IL	30	69	184	4,797,273	69,414
IN	30	65	166	2,238,147	34,630
KS	29	63	186	1,000,398	15,786
KY	29	64	196	1,120,575	17,379
LA	30	63	146	1,201,510	19,111
MA	31	69	177	10,695,244	154,891
MD	28	69	190	2,491,372	36,154
ME	32	72	187	1,215,042	16,787
MI	30	67	173	3,257,719	48,499
MN	30	68	176	1,768,685	25,937
MO	29	67	191	1,937,271	28,953
MS	31	66	188	514,513	7,751
MT	33	69	169	305,490	4,424
NC	29	66	190	2,507,978	37,811
ND	29	65	149	220,194	3,387
NE	30	69	192	536,243	7,798
NH	32	81	277	1,873,150	23,081
NJ	32	73	213	7,459,863	102,554
NM	29	65	194	496,633	7,691
NV	41	93	238	1,241,557	13,377
NY	33	76	201	11,675,730	152,723
OH	29	66	176	3,843,913	57,932
OK	29	67	179	877,034	13,084
OR	28	64	187	1,278,495	19,997
PA	31	69	188	4,928,156	71,159
RI	31	70	176	1,157,634	16,559
SC	28	68	217	1,232,020	18,117
SD	34	74	217	247,896	3,349
TN	30	66	167	1,754,011	26,551
TX	26	62	184	6,154,744	98,592
UT	26	67	240	615,343	9,173
VA	28	68	201	3,038,906	44,872
VT	31	72	182	325,751	4,526
WA	27	66	195	1,548,626	23,514
WI	30	68	210	2,268,723	33,298
WV	32	68	177	393,698	5,779
WY	30	68	185	175,653	2,567
Total	30	70	195	121,853,167	1,751,067

Table B.2: **Summary statistics across residence locations - Frequent Users**

variable	min	p10	p25	p50	p75	p90	max	mean	N
Sales locations visited	1	5	9	13	20	27	139	15.29	1,729
Sales locations within 120km	7	80	137	231	494	838	1,111	344.2	1,729
Share available loc. visited	0	0.02	0.03	0.05	0.08	0.13	0.43	0.07	1,729

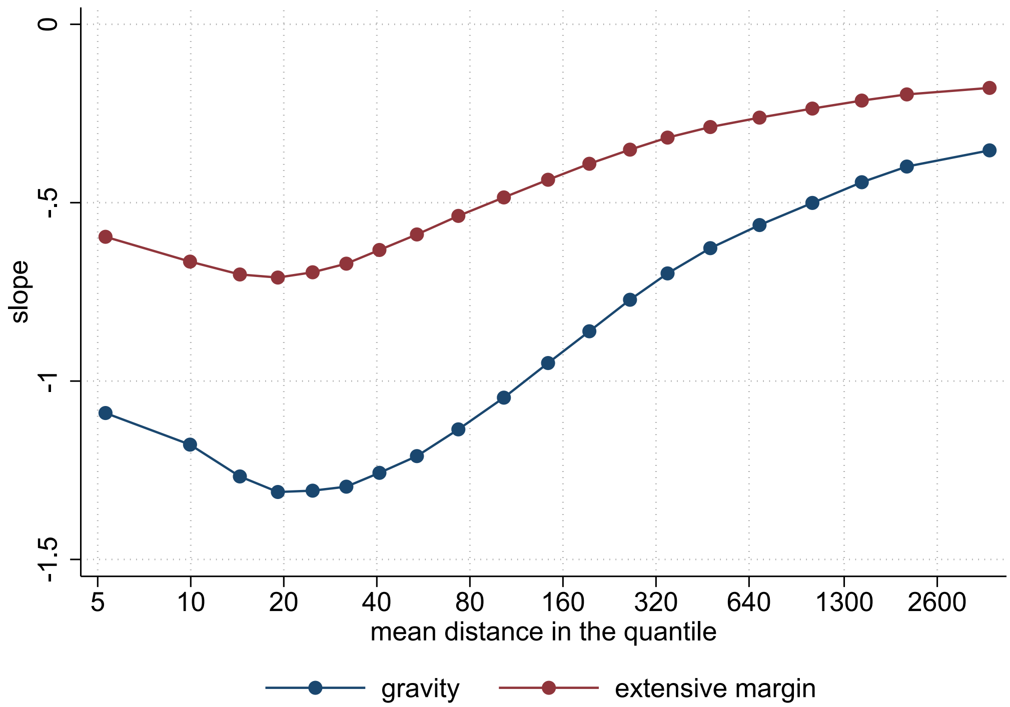

Figure B.1: Gravity in Expenditure

B.5 Margins decomposition

Figure B.2 shows the margins decomposition in eq. (4) applied to estimates of the gravity regression (2). The accounts margin is associated to 57% of the decline in expenditure in the typical sector as distance increases.

Tables B.5 shows the actual values of the account and expenditure margin with associated p-values represented in Figure 2; Table B.6 shows the actual values of the account and expenditure margin with associated p-values represented in Figure B.2

Tables B.7 and B.8 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.

Table B.3: **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	327.7	1,600.3	4,130	8,237.4	454.7
Apparel	0	4.7	15.6	52.1	364.7	3,833.6	8,253.1	201.4
Auto Repair/Service/Parking	0	0	7.9	24	78	2,315.3	7,937.3	94.8
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	1,493.2	7,868.1	49.6
Communications	0	6.5	24.3	685.1	2,018	3,944.5	8,134.9	551.5
Durable Goods	0	5.7	21.8	161.8	1,650.7	3,946.5	7,115	419.5
Eating and Drinking Places	0	0	12.6	50.4	496.4	3,739.9	8,254.8	217.2
Educational Services	0	4.5	21.8	141.1	957.9	4,024.4	7,981.4	303.3
Food Stores	0	0	4.2	15.3	54	2,416.2	8,218	93.8
Furniture, Home Furnishings, Equip.	0	1.8	11.2	26.3	133.6	3,299.6	8,243.6	136.8
Gasoline Services	0	0	8.9	34.6	275.1	2,278.2	8,233.3	126.9
General Merchandise Stores	0	0	8.7	20.8	61.4	2,001.5	8,223.9	87.4
Health Services	0	0	8.6	20.3	46.3	2,231.1	7,969.9	83.9
Hospitality	51.3	162.8	367	1,011.8	2,257.9	4,158.5	8,253.1	801.9
Misc. Retail	0	0	8.6	29.6	355.1	3,736.8	8,223.9	193.1
Misc. Services	0	2.2	15.8	67.8	1,131.8	3,905.3	7,765.3	302.2
Motion Pictures	0	0	5.7	16.7	64.4	3,756.9	7,884.2	126.5
NonDurable Goods	0	0	8.2	22	143.9	3,437.6	7,768.4	145.4
Other Vehicles Sales/Service/Parts	0	6.6	20.1	55.1	506.1	3,018.8	7,879.4	191.5
Personal Services	0	0	6.7	20	135.1	3,332.5	8,251.4	132.3
Transportation Services	0	14.1	34.7	212.8	1,502.2	4,190.6	8,158.5	408.1
Vehicle Rental	0	6.8	79.9	1,506	2,813.7	4,558.8	8,216.8	893.7
Total	0	0	9.2	30.3	293.2	3,355.2	8,254.8	163.6

Table B.4: **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	8.3	37.5	418.8	1,752.7	4,290.3	8,237.4	530.2
Apparel	0	5.3	16.8	56.1	437.4	3,864.6	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	1,577.2	7,868.1	56.3
Communications	0	4.5	14.7	113.6	1,518.2	3,818.2	8,134.9	367.8
Durable Goods	0	10.5	30.5	197.9	1,866	4,017.6	7,115	454.3
Eating and Drinking Places	0	0.8	15.4	79.2	708.8	3,940.4	8,254.8	264.2
Educational Services	0	8.6	24.6	107.8	632	3,913.2	7,981.4	262.4
Food Stores	0	0	5.2	16.9	55	2,378.1	8,218	91.9
Furniture, Home Furnishings, Equip.	0	4.6	13.1	30.7	129.7	2,970.8	8,243.6	129.9
Gasoline Services	0	0	9.7	39.6	320.1	2,252.9	8,233.3	133.9
General Merchandise Stores	0	0	9.9	23.1	77.2	2,555.2	8,223.9	104.1
Health Services	0	0	9.8	24.9	75.7	2,658.9	7,969.9	112.2
Hospitality	59.5	179.2	434.7	1,320.4	2,665	4,331.4	8,253.1	949.3
Misc. Retail	0	0	13	49.7	704.6	3,911.7	8,223.9	254.8
Misc. Services	0	5.3	17.2	54.7	666.7	3,964.9	7,765.3	238.8
Motion Pictures	0	0	7.3	22.2	223.1	3,960.8	7,884.2	181.9
NonDurable Goods	0	3	11.2	34.1	742.2	3,942.2	7,768.4	249.7
Other Vehicles Sales/Service/Parts	0	9.1	23.1	64.9	939	3,139	7,879.4	244.9
Personal Services	0	0	11	38.5	526.8	3,856.4	8,251.4	218.3
Transportation Services	4	16.9	77	651.5	2,004.8	7,346.6	8,158.5	683.3
Vehicle Rental	0	10.5	509.5	1,789.3	3,280.8	4,813.8	8,216.8	1,099.8
Total	0	0	12.8	44	521.3	3,822.1	8,254.8	224

Table B.5: **Expenditure out of home place (distances up to 120km)**

Category	Overall coeff	Overall pv	Accounts Margin coeff	Accounts Margin pv	Expenditure Margin coeff	Expenditure Margin pv	Share Accounts Margin	Obs.
Food Stores	-2.23	0.00	-1.12	0.00	-1.11	0.00	0.50	22,652
Gasoline Services	-2.08	0.00	-0.97	0.00	-1.11	0.00	0.47	39,673
General Merchandise Stores	-1.79	0.00	-1.08	0.00	-0.71	0.00	0.60	26,845
Misc. Retail	-1.70	0.00	-1.07	0.00	-0.63	0.00	0.63	34,057
Eating and Drinking Places	-1.57	0.00	-0.93	0.00	-0.64	0.00	0.59	34,509
Building Mat./Hardware/Garden Supp.	-1.40	0.00	-0.87	0.00	-0.53	0.00	0.62	14,190
Auto Repair/Service/Parking	-1.25	0.00	-0.88	0.00	-0.38	0.00	0.70	4,415
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	-0.36	0.00	0.68	5,136
Apparel	-1.10	0.00	-0.83	0.00	-0.27	0.00	0.75	15,921
Transportation Services	-1.09	0.00	-0.68	0.00	-0.41	0.07	0.62	743
Furniture, Home Furnishings, Equip.	-1.07	0.00	-0.85	0.00	-0.23	0.00	0.79	12,292
Auto and Truck Sales/Service/Parts	-1.04	0.00	-0.81	0.00	-0.23	0.00	0.78	7,302
Motion Pictures	-1.04	0.00	-0.85	0.00	-0.19	0.00	0.82	1,927
Amusement, Rec. Serv.	-1.02	0.00	-0.66	0.00	-0.36	0.00	0.64	2,959
Educational Services	-1.00	0.00	-0.86	0.00	-0.13	0.56	0.87	712
Personal Services	-0.96	0.00	-0.89	0.00	-0.07	0.12	0.93	5,204
Vehicle Rental	-0.95	0.00	-0.83	0.00	-0.11	0.50	0.88	546
Misc. Services	-0.92	0.06	-0.63	0.00	-0.29	0.52	0.69	222
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	0.11	0.76	552
Other Vehicles Sales/Service/Parts	-0.68	0.41	-0.71	0.00	0.03	0.97	1.04	257
Hospitality	-0.65	0.01	-0.49	0.00	-0.17	0.35	0.75	1,394
Durable Goods	-0.09	0.90	-0.27	0.04	0.18	0.76	3.15	79

Table B.6: **Gravity in expenditure (distances up to 120km)**

Category	Overall coeff	Overall pv	Accounts Margin coeff	Accounts Margin pv	Expenditure Margin coeff	Expenditure Margin pv	Share Accounts Margin	Obs.
Food Stores	-0.85	0.00	-0.36	0.00	-0.50	0.00	0.42	18,635
Gasoline Services	-0.60	0.00	-0.25	0.00	-0.35	0.00	0.41	34,621
General Merchandise Stores	-0.93	0.00	-0.50	0.00	-0.43	0.00	0.54	23,933
Misc. Retail	-0.65	0.00	-0.40	0.00	-0.25	0.00	0.61	30,046
Eating and Drinking Places	-0.56	0.00	-0.31	0.00	-0.25	0.00	0.55	31,028
Building Mat./Hardware/Garden Supp.	-0.73	0.00	-0.39	0.00	-0.34	0.00	0.53	11,610
Auto Repair/Service/Parking	-0.40	0.00	-0.23	0.00	-0.16	0.00	0.59	3,014
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	-0.08	0.09	0.75	3,914
Apparel	-0.53	0.00	-0.36	0.00	-0.17	0.00	0.67	14,069
Transportation Services	-0.47	0.00	-0.16	0.00	-0.31	0.00	0.34	635
Furniture, Home Furnishings, Equip.	-0.57	0.00	-0.40	0.00	-0.17	0.00	0.70	10,740
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	0.24	0.81	1,253
Amusement, Rec. Serv.	-0.22	0.00	-0.10	0.00	-0.12	0.00	0.45	2,330
Educational Services	-0.15	0.38	-0.19	0.00	0.04	0.83	1.24	530
Personal Services	-0.31	0.00	-0.27	0.00	-0.04	0.26	0.86	3,761
Vehicle Rental	-0.08	0.59	-0.22	0.00	0.14	0.30	2.71	296
Misc. Services	0.97	0.01	-0.10	0.07	1.07	0.00	-0.10	120
Communications	-0.41	0.01	-0.25	0.00	-0.15	0.22	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	0.38	0.55	1,160
Durable Goods	1.11	0.67	0.00	1.00	1.11	0.67	0.00	15

Table B.7: **Expenditure out of home place: number of transactions and average expenditure**

Category	Expenditure margin coeff	pv	Batch size margin coeff	pv	Frequency margin coeff	pv	Share of Frequency margin	Share of Account+Frequency margins	Obs.
Food Stores	-1.11	0.00	-0.18	0.00	-0.93	0.00	0.84	0.92	22,652
Gasoline Services	-1.11	0.00	-0.09	0.00	-1.02	0.00	0.92	0.96	39,673
General Merchandise Stores	-0.71	0.00	-0.06	0.00	-0.65	0.00	0.91	0.97	26,845
Misc. Retail	-0.63	0.00	0.05	0.00	-0.68	0.00	1.08	1.03	34,057
Eating and Drinking Places	-0.64	0.00	0.02	0.05	-0.66	0.00	1.04	1.02	34,509
Building Mat./Hardware/Garden Supp.	-0.53	0.00	-0.02	0.45	-0.51	0.00	0.96	0.99	14,190
Auto Repair/Service/Parking	-0.38	0.00	-0.21	0.00	-0.16	0.00	0.44	0.83	4,415
NonDurable Goods	-0.11	0.45	0.02	0.88	-0.13	0.04	1.17	1.02	978
Health Services	-0.36	0.00	-0.17	0.00	-0.19	0.00	0.52	0.85	5,136
Apparel	-0.27	0.00	-0.01	0.53	-0.26	0.00	0.95	0.99	15,921
Transportation Services	-0.41	0.07	-0.26	0.16	-0.15	0.19	0.37	0.76	743
Furniture, Home Furnishings, Equip.	-0.23	0.00	-0.01	0.88	-0.22	0.00	0.97	0.99	12,292
Auto and Truck Sales/Service/Parts	-0.23	0.00	-0.01	0.85	-0.22	0.00	0.96	0.99	7,302
Motion Pictures	-0.19	0.00	0.02	0.69	-0.21	0.00	1.11	1.02	1,927
Amusement, Rec. Serv.	-0.36	0.00	-0.19	0.01	-0.18	0.00	0.49	0.82	2,959
Educational Services	-0.13	0.56	-0.11	0.63	-0.03	0.69	0.22	0.89	712
Personal Services	-0.07	0.12	0.16	0.00	-0.23	0.00	3.31	1.17	5,204
Vehicle Rental	-0.11	0.50	-0.05	0.75	-0.06	0.29	0.56	0.95	546
Misc. Services	-0.29	0.52	-0.20	0.65	-0.09	0.36	0.32	0.79	222
Communications	-0.28	0.04	-0.17	0.25	-0.11	0.09	0.38	0.81	424
Agricultural Services	-0.21	0.11	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	-0.17	0.35	-0.05	0.77	-0.12	0.09	0.72	0.93	1,394
Durable Goods	0.18	0.76	0.02	0.97	0.17	0.50	0.91	1.19	79

Table B.8: **Gravity in expenditure: number of transactions and average expenditure**

Category	Expenditure margin coeff	pv	Batch size margin coeff	pv	Frequency margin coeff	pv	Share of Frequency margin	Share of of Account+Frequency margins	Obs.
Food Stores	-0.50	0.00	-0.13	0.00	-0.36	0.00	0.73	0.84	18,635
Gasoline Services	-0.35	0.00	-0.04	0.00	-0.31	0.00	0.89	0.93	34,621
General Merchandise Stores	-0.43	0.00	-0.09	0.00	-0.33	0.00	0.78	0.90	23,933
Misc. Retail	-0.25	0.00	-0.01	0.16	-0.24	0.00	0.95	0.98	30,046
Eating and Drinking Places	-0.25	0.00	-0.02	0.00	-0.23	0.00	0.90	0.96	31,028
Building Mat./Hardware/Garden Supp.	-0.34	0.00	-0.07	0.00	-0.27	0.00	0.80	0.91	11,610
Auto Repair/Service/Parking	-0.16	0.00	-0.09	0.06	-0.07	0.00	0.43	0.77	3,014
NonDurable Goods	-0.24	0.01	-0.09	0.23	-0.15	0.00	0.62	0.86	758
Health Services	-0.08	0.09	0.03	0.53	-0.11	0.00	1.33	1.08	3,914
Apparel	-0.17	0.00	-0.02	0.11	-0.15	0.00	0.90	0.97	14,069
Transportation Services	-0.31	0.00	-0.09	0.24	-0.22	0.00	0.70	0.80	635
Furniture, Home Furnishings, Equip.	-0.17	0.00	-0.04	0.06	-0.13	0.00	0.77	0.93	10,740
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	0.24	0.02	0.64	-0.09	0.02	1.32	1.06	1,253
Amusement, Rec. Serv.	-0.12	0.00	-0.04	0.33	-0.08	0.00	0.67	0.82	2,330
Educational Services	0.04	0.83	-0.01	0.94	0.05	0.37	1.33	0.92	530
Personal Services	-0.04	0.26	0.08	0.02	-0.12	0.00	2.79	1.25	3,761
Vehicle Rental	0.14	0.30	0.19	0.16	-0.05	0.28	-0.35	3.31	296
Misc. Services	1.07	0.00	1.19	0.00	-0.12	0.14	-0.12	-0.23	120
Communications	-0.15	0.22	-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	0.38	-0.05	0.39	-0.01	0.75	0.16	0.62	1,160
Durable Goods	1.11	0.67	0.96	0.73	0.15	0.79	0.14	0.14	15

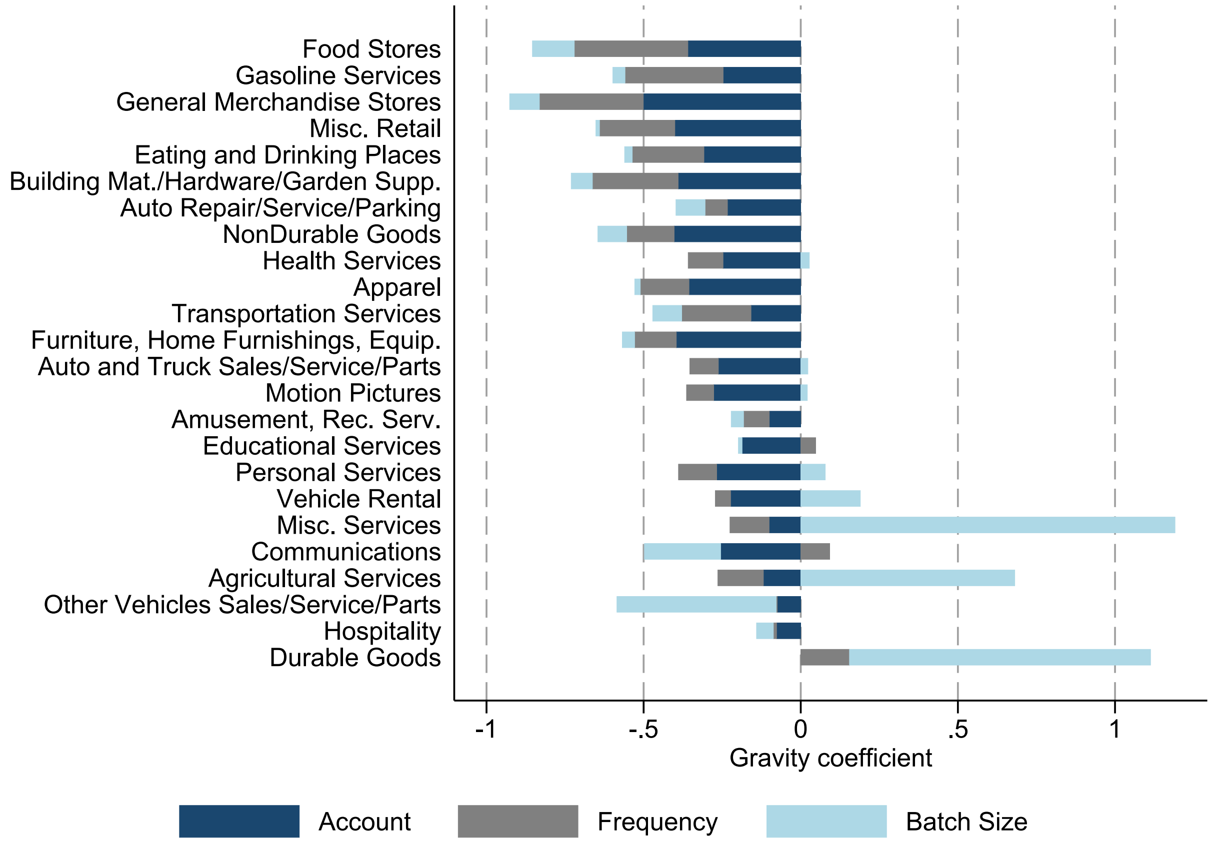

Figure B.2: Margins in gravity regression

B.6 Gravity and the frequency of transactions

These figures show further robustness on the relation between gravity and the frequency of transactions. Figure B.3 shows the correspondent of Figure 3 using all coefficients, and not just the ones significantly different from zero. Figure B.4 use the strength of gravity as measured by regression (2), only showing the coefficients significantly different from zero. Figure B.5 shows the correspondent of figure B.4, using all estimated slopes.

B.7 Rain and the flattening of gravity

Figure B.6 below parallels figure 4 in the main text, and includes all coefficients where the interaction between out-of-home expenditure and rain is significantly different from zero.

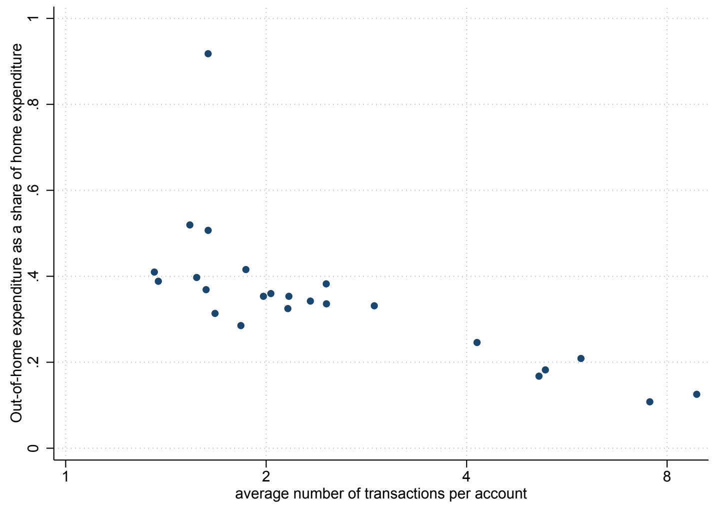

Figure B.3: Drop in expenditure out of home (all coefficients)

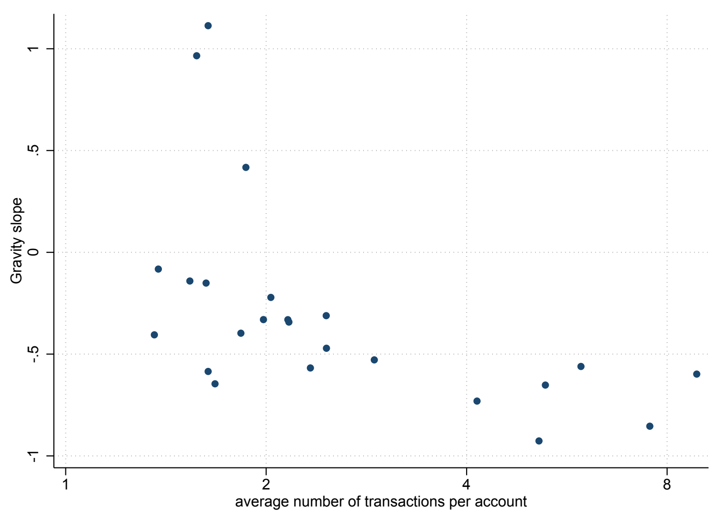

Figure B.4: Gravity and frequency of transactions (only significant slopes)

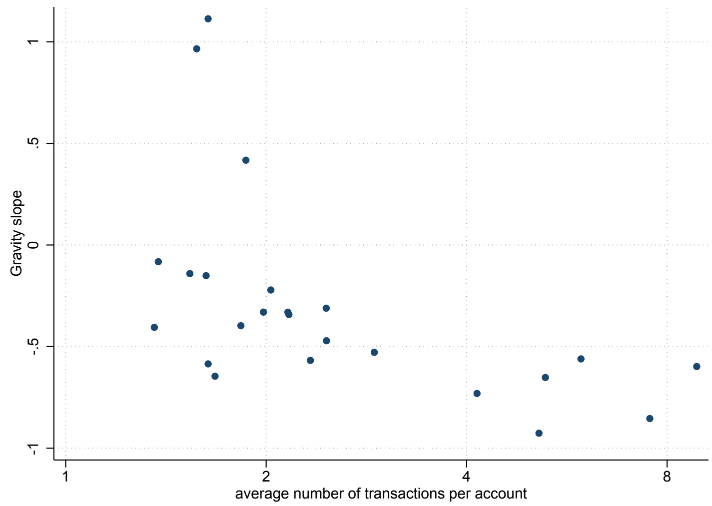

Figure B.5: Gravity and frequency of transactions (all slopes)

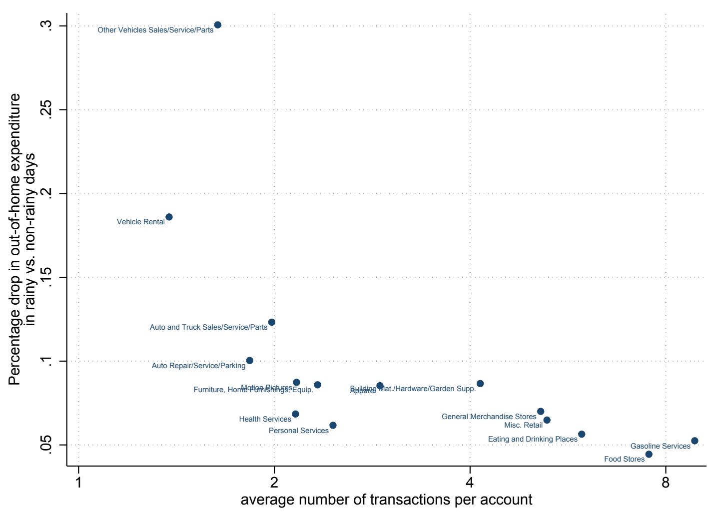

Figure B.6: The flattening in gravity across sectors