DISCUSSION PAPER SERIES

DP12150
(v. 3)

CONSUMER MOBILITY AND THE LOCAL STRUCTURE OF CONSUMPTION INDUSTRIES

Ferdinando Monte, J Bradford Jensen and Sumit Agarwal

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

ISSN 0265-8003

CONSUMER MOBILITY AND THE LOCAL STRUCTURE OF CONSUMPTION INDUSTRIES

Ferdinando Monte, J Bradford Jensen and Sumit Agarwal

Discussion Paper DP12150
First Published 15 July 2017
This Revision 27 January 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

CONSUMER MOBILITY AND THE LOCAL STRUCTURE OF CONSUMPTION INDUSTRIES

Abstract

We study local employment, establishment density, and establishment size across industries delivering final consumption, which comprise a substantial fraction of production, shape local amenities, and pay different wages. In a stylized model of consumer mobility, lower industry storability/durability concentrates demand in space, increasing equilibrium employment. Credit card transactions data show that consumer mobility is limited and varies substantially across sectors; moreover, expenditure declines more rapidly with distance in sectors transacted more frequently. Lower storability/durability, proxied by average transaction frequency, increases a sector's local employment via higher establishment density. Variation in consumer mobility is as economically significant as consumers' expenditure shares.

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

Keywords: N/A

Ferdinando Monte - ferdinando.monte@georgetown.edu
Georgetown University MSB and CEPR

J Bradford Jensen - jbj24@georgetown.edu
Georgetown University MSB and Peterson Institute

Sumit Agarwal - ushakri@yahoo.com
National University of Singapore

Acknowledgements
We would like to thank David Atkin, Andrew Bernard, Victor Couture, Jonathan Dingel, Colin Hottman, Nate Miller, Steve Redding, Andrés Rodriguez-Clare, Esteban Rossi-Hansberg, Tony Yezer, and seminar participants at Columbia, Harvard-MIT, National University of Singapore, NBER, Philadelphia and Richmond FED, Singapore Management University, Southern Methodist University and University of Virginia for helpful comments and discussions.

Consumer Mobility and the Local Structure of Consumption Industries*

Sumit Agarwal† J. Bradford Jensen‡ Ferdinando Monte§

January 2020

(please click here for the latest version)

Abstract

We study local employment, establishment density, and establishment size across industries delivering final consumption, which comprise a substantial fraction of production, shape local amenities, and pay different wages. In a stylized model of consumer mobility, lower industry storability/durability concentrates demand in space, increasing equilibrium employment. Credit card transactions data show that consumer mobility is limited and varies substantially across sectors; moreover, expenditure declines more rapidly with distance in sectors transacted more frequently. Lower storability/durability, proxied by average transaction frequency, increases a sector's local employment via higher establishment density. Variation in consumer mobility is as economically significant as consumers' expenditure shares.

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

*A previous version of this paper circulated with the title "The Geography of Consumption". Sumit Agarwal: ushakri@yahoo.com; Brad Jensen: jbj24@georgetown.edu; Ferdinando Monte: ferdinando.monte@georgetown.edu. We would like to thank David Atkin, Andrew Bernard, Victor Couture, Jonathan Dingel, Colin Hottman, Nate Miller, Steve Redding, Andrés Rodriguez-Clare, Esteban Rossi-Hansberg, Tony Yezer, and seminar participants at Columbia, Harvard-MIT, National University of Singapore, NBER, Philadelphia and Richmond FED, Singapore Management University, Southern Methodist University and University of Virginia for helpful comments and discussions.

†National University of Singapore, Singapore.

‡Georgetown University, McDonough School of Business, Washington, D.C.; and, NBER.

§Georgetown University, McDonough School of Business, Washington, D.C.; NBER; and, CEPR.

1 Introduction

The delivery of final consumption involves around forty percent of employment and twenty-five percent of GDP in modern economies. These large magnitudes are consequential for the fortunes of small towns and big cities alike. The local composition of these industries, like apparel stores, gasoline stations, health services, or restaurants, helps to shape "amenities," a component of the residential attractiveness of different locations, by modifying consumption opportunities. Amenities further affect housing values, the spatial sorting of skilled individuals, and wage and well-being inequality.[1] These industries also pay different wages at the national level, in part reflecting different productivity and skill demand: for example, monthly wages in the United States in 2007 for restaurant employees averaged one thousand dollars but reached well beyond three thousand dollars for workers in health services.[2]

Despite their potentially far-reaching implications, surprisingly little is known about the market incentives that shape the relative supply of these different industries. We investigate the empirical determinants of the "local structure of consumption industries," i.e., the local composition of employment, number of establishments, and establishment size across industries that deliver final consumption.

A natural starting point for our discussion is that, in industries delivering final goods and services, firms make production and location decisions taking into account a consumer's willingness to travel to the point of sale. Households devote considerable resources to procuring final consumption: data from the 2017 National Household Transportation Survey, for example, show that, among all trips by privately-owned vehicles, more than 67% were taken for purposes like shopping and errands or recreation. Individual mobility is, largely, consumer mobility. We then organize our analysis around a highly stylized framework where consumers solve an optimal travel distance and inventory problem. Their choices are a function of individual travel costs and the storability/durability of a particular sector's output. Heterogeneity in individual travel costs generates a demand that smoothly decays with distance. In lower storability/durability sectors, agents optimally choose lower inventories and a higher frequency of purchase trips for a given total consumption; hence, consumers' willingness to travel long distances per trip falls and their demand decays faster with distance. Producers, who use land and labor, respond by limiting the use of the locations most distant from consumers, and substitute this input with workers; in equilibrium, when storability/durability is lower, local employment is higher.

We examine individual purchase trips in more than 1.7 million credit card transactions by 70 thousand American consumers in 2003, to provide the first cross-sectoral description of consumers' local geographical mobility (more simply, "consumer mobility"). Consumer mobility is quite limited: agents appear to purchase from just a few among the geographically proximate locations, and the median transaction occurs at a distance of about 9 km from the home location. However, distances traveled and the associated decline in expenditure, or "gravity," vary substantially across sectors: for example, expenditure out of the town of residence of consumers ("out-of-home") is about 33% of the expenditure in the home town of residence ("at-home") for apparel stores, but only 11% for food stores. Most of the sectoral variation in gravity is accounted for by extensive margins of adjustment, like individuals taking fewer out-of-home

[1] See for example Diamond (2016) and Couture, Gaubert, Handbury, and Hurst (2019).

[2] This computation uses the County Business Patterns 2007 and the set of industries described below.

trips, or by fewer individuals traveling out-of-home altogether.

Importantly, the data clearly show that gravity is stronger in sectors where transactions are more frequent on average. This fact is a direct implication of our stylized framework: lower storability/durability implies both a higher average frequency of transactions and a faster decline of expenditure with distance. Note that our simple model suggests that the average frequency of transactions is a proxy for durability/storability, after controlling for consumers' expenditure shares across sectors. We construct consumers' budget shares across sectors using the CEX expenditure survey from the U.S. Bureau of Labor Statistics in 2003. The negative correlation between gravity and frequency is robust to controlling for expenditure shares and is then consistent with consumer mobility decisions influencing the local structure of consumption industries. We use the frequency of transactions as a proxy for durability/storability in the rest of the paper and control for CEX expenditure share throughout.

Of course, one must interpret cross-sectoral correlations with caution. It is in principle possible that consumers do not choose distances traveled as a function of storability/durability, and that the observed sectoral variation in consumer mobility merely reflects differences in establishment density due to supply-side industry characteristics, like fixed costs. In supplemental results, reassuringly, we report evidence consistent with a specific role for consumers' decisions in distances traveled. The short span of our credit card data, about eight months, allows us to reasonably assume a fixed supply network. We examine the role of individual agents' income on the geographical distribution of their trips, as proxied with the fraction of out-of-home transactions. We also compare the travel behavior of the same individual under two travel costs regimes as a function of that individual's income. In particular, we assemble daily precipitation data from thousands of weather stations across the United States; we reconstruct local weather conditions at the time of each purchase; we then examine the difference in spatial distribution of transactions of an individual during rainy vs. non-rainy days, again as a function of income. In both cases, we control for many potentially confounding factors with the use of available data or fixed effects. We find that agents' income modulates the spatial distribution of purchases in a way that is related to the sector average frequency of transactions. Since the supply side is fixed, differences in mobility across sectors that are related to individual income are likely to reflect heterogeneous individual decisions.

To quantify the influence of consumer mobility on the local structure of consumption industries, we examine how sectors with different storability/durability respond to variation in population size. This exercise leverages an implication of consumers' more limited willingness to travel for low storability/durability sectors: a larger local population should imply a relatively more substantial increase in local employment in high-frequency sectors, relative to low-frequency sectors. We evaluate this implication using County Business Patterns data from 1998 and 2007 (dates around our credit card sample period). We address the endogeneity problem arising from regressing county-sector level employment on the county population relying on geological variation in the presence of aquifers.[3]

We find that a larger county population induces a more substantial increase in employment in high-frequency sectors relative to low-frequency sectors. This increase is wholly accounted for by an increase in the number of establishments, which is consistent with a reduction in the average distance between

[3] We follow intuitions in Burchfield, Overman, Puga and Turner (2006) and Duranton and Turner (2017).

consumers and points of sale within the county. We find that the economic significance of the consumer mobility channel is as big as the economic significance of the variation associated with consumers' expenditure shares. In ancillary results, we also show evidence that the spillover of a larger local population on neighboring counties' employment via consumer mobility is relatively smaller in high-frequency sectors, as a more geographically concentrated local demand would imply; however, conditional on the frequency of transactions, this spillover is still relatively stronger in sectors with larger consumer expenditure shares. Several robustness exercises further limit concerns about confounding factors.

This work contributes to the literature in three related ways.

Our main contribution is to analyze, in a systematic way across sectors, the determinants of the local structure of consumption industries. We find evidence that a lower storability/durability of a sector's output increases its local employment and local establishment density, consistent with a central role of consumer mobility in shaping local equilibrium outcomes. This contribution is significant in two directions. First, the empirical determinants of the structure of these industries have received little systematic attention in the literature.[4] Yet, they are likely consequential for the fortunes of different localities. Industries that cater to final demand account for a major share of employment and GDP. Their amenity value is prominently associated with the attractiveness of an area; this attractiveness varies for different demographic and skill groups, with consequences on local productivity, house prices, wages and well-being inequality.[5] Further, data from the U.S. County Business Patterns shows that in 2007, the standard deviation of average monthly wages across these industries was more than twelve hundred U.S. dollars, plausibly reflecting considerable variation in skill demand and productivity, and further influencing the desirability of particular regions. Second, our results complement a more traditional view of the sectoral composition of locations based on Central Place Theory, where sectors with larger overall expenditure (or smaller economies of scale, or the more commonly demanded varieties within a sector) are found in most cities, while sectors with the smaller expenditure (or most significant economies of scale, or the less commonly demanded varieties within a sector) are observed only in large cities.[6] Our results indicate that after controlling for consumers' budget shares, consumer mobility appears to have an independent and economically equal role.

Our second contribution is to provide insight into the nature of the spatial frictions that influence consumer behavior: lower storability/durability of a sector's output reduces consumers' optimal travel length decisions, and generates a faster decline of expenditure with distance. This contribution is important for two reasons. First, it complements and brings together a growing literature related to spatial frictions in consumption markets. This literature mostly focuses on specific sectors[7] and as such it is

[4]The vast literature about the sectoral composition of locations has focused on broad sectors, narrower sectors within manufacturing, or occupational functions: see, for example, Duranton and Puga (2001, 2004, 2005), Rosenthal and Strange (2004), Ellison, Glaeser and Kerr (2010), Davis and Dingel (2019).

[5]See for example Glaeser, Kolko, and Saiz (2000), Diamond (2016), Couture and Handbury (2019), Couture, Gaubert, Handbury, and Hurst (2019). The work on endogenous consumption amenities has mostly taken a homogeneous view of these different consumption sectors and focused on the role of income.

[6]See Christaller (1933) for foundations. For theoretical work on Central Place Theory, see for example Hsu (2012) or Hsu, Holmes, and Morgan (2014); for an empirical test, see Schiff (2015).

[7]The literature has for example considered food distribution (Allcott et al., 2018), gasoline (Houde, 2012), health care (Raval and Rosenbaum, 2018), movie theaters (Davis, 2006), or restaurants (Couture, 2016, Davis, Dingel, Monras and Morales, 2018). Some work considers the response of cross-border consumer travel in response to exchange rates (Chandra,

able to accommodate sector-specific considerations. Our unified view necessarily abstracts from these specificities to highlight a common feature, storability/durability, that appears to be economically significant in explaining the local structure of consumption industries.[8] Second, this contribution speaks to the recent literature on the tradeability of services,[9] which has mostly focused on estimating the decline of service flows with distance, with less focus on the underlying reasons why distance affects trade. Our results indicate that, for the significant fraction of economic activity related to consumer behavior, a lower storability/durability of a sector's output makes it appear less "tradeable."[10] Understanding the nature of these frictions is important for the effect that local policies, like new transportation infrastructure, may have across sectors. Relatedly, our results also suggest that, everything else equal, innovative on-line technologies to deliver consumption goods might find more natural targets in markets where consumer mobility frictions matter most.[11]

Our third contribution is to quantify salient features of consumer mobility in a systematic way across industries for the first time. We establish new sector-level facts like the distributions of traveled distances, gravity estimates, and the margins of expenditure decline. A significant fraction of the extensive literature in quantitative spatial economics[12] needs to make assumptions about and calibrate small-geography models. Our analysis is the first to provide direct evidence, comparable across sectors, to inform those choices. This evidence ultimately allows for improved positive and normative quantitative implications of those studies.

The rest of the paper is organized as follows. Section 2 introduces a highly stylized model to describe our main intuition about consumer mobility and local equilibrium outcomes. In Section 3, we present a set of cross-sectoral stylized facts on consumer mobility. Section 4 investigates empirically the local structure of consumption industries. Section 5 describes ancillary results, robustness analyses and limitations of our work. Section 6 concludes.

Head, and Tappata, 2014), or tax rates (Agarwal, Marwell and McGranahan, 2017; Baker, Johnson and Kueng 2018).

[8] Previous work has shown that durability/storability is important for the study of commodity markets (e.g., Williams and Wright, 1991, Coleman, 2009); obtaining consistent estimates of demand elasticity (Hendel and Nevo 2006a,b); the consequences of exchange rate devaluations (Alessandria, Kaboski and Midrigan, 2010); understanding retail pricing strategies and measuring inflation (Chevalier and Kashyap 2018); accounting for consumer response to local sales taxes (Baker, Johnson and Kueng, 2018).

[9] See for example Jensen and Kletzer (2010), Gervais and Jensen (2019), or Eckert (2019).

[10] These considerations are reminiscent of intuitions that can be traced back to Von Thunen's (1936) model of rural land use, or Karádi and Koren (2017) for a more modern version; for an empirical test of the Von Thunen model, see for example Fafchamps and Shilpi (2003) about Nepal.

[11] While online transactions are inherently less suitable for some consumption activities (like a restaurant dinner), our results speak to frictions that makes some industries natural targets for e-commerce. On-line transactions grew from about 0.9% of total retail sales in 2000 to 6.4% in 2014 (Hortaçsu and Syverson, 2015), arguably to ease the burden of travel frictions that we are analyzing here. The literature about on-line transactions is growing and includes for example papers on the role of taxes in determining sales of on-line versus traditional retailers (Ellison and Ellison, 2009), the importance of distance and the persistence of home bias in on-line auctions (Hortaçsu, Martínez-Jerez and Douglas, 2009), or the gains from e-commerce (Dolfen et al., 2019).

[12] See Redding and Rossi-Hansberg (2017) for a survey.

2 A Simple Model of Shopping

We start by introducing a highly stylized model of consumer behavior and equilibrium employment to frame our discussion. Our setting is essentially a mirror image of an Alonso-Mills-Muth monocentric city model: all agents live in a single place, and they choose where to shop. Since they want to consume at a constant rate over time, but travel is costly, an inventory problem emerges. We generalize ideas present in Oi (1992) – but dating back at least to Baumol (1952) – to a setting where consumers with heterogeneous travel costs choose 1) how far to travel for their purchases, 2) how frequently to do so, and 3) the purchase size per trip. Producers use homogeneous labor and land heterogeneous in its distance j to consumers, to supply a good only differentiated by its point of sale. A given price function $p(j)$ induces a spatial distribution of supply from optimal producers' choices; and a spatial distribution of demand, from optimal consumers' choices. The equilibrium price function makes demand and supply identical point-wise and determines the marginal plot of land used. The good is characterized by a storability/durability parameter ("storage cost" for brevity) g. We think of storability/durability as a general characteristic of the sector, capturing the length of time over which a good or service can deliver its utility flow. For perishable items, this concept is intuitive: fruit, for example, will depreciate if not eaten quickly. More generally, this concept may reflect depreciation due to use (as in the case of a shirt) or to the passage of time (as in the case of a haircut). This parameter g regulates, in equilibrium, the spatial decay in demand and hence affects the equilibrium distribution of supply and total employment.

2.1 Producers

There is one sector with productivity A. Producers operate in perfect competition and are potentially active in any location $j \in [0, +\infty)$. Each location j is endowed with a fixed amount of land $\bar{D}$. A firm located in j uses land and labor $L(j)$ to produce goods with the production function $Q(j) = A\bar{D}^{1-\beta}L(j)^{\beta}$. Firms located in j choose labor to maximize profits:

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

where $p(j)$ and $R(j)$ are the output price and the rental rate of land at j; w is the wage, which we assume fixed and determined in an outside sector[13]. The optimal quantity of labor is given by

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

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

$$Q(j) = A^{1/(1-\beta)}\bar{D}\left(\frac{\beta}{w}\right)^{\beta/(1-\beta)} p(j)^{\beta/(1-\beta)} \tag{2}$$

[13] We generalize this assumption below.

Absentee landlords rent their land to the highest price. Under free entry, positive profits in a location bid up these land prices; in equilibrium, the price of land is $R(j) = \bar{R} \cdot p(j)^{1/(1-\beta)}$ and profits are zero everywhere.[14]

2.2 Consumers

A measure N of consumers is heterogeneous in t, an increasing index of individual travel costs. All consumers (exogenously) live in location 0. Each agent wants to consume a fixed quantity $\bar{q}$ of the good over one unit of time: in particular, in a fraction di of the unit time, the consumer eats a fixed quantity $\bar{q} \times di$ of the good. Assuming a fixed quantity consumed simplifies the problem and emphasizes the role of the price function $p(j)$ in allocating consumers to purchase locations.

A consumer with travel cost t wants to minimize the total cost of consuming $\bar{q}$ units of the good per unit of time, $c(t)$.[15] In particular, the consumer takes as given $p(j)$ and solves:

$$c(t) = \min_{j,z} p(j)\,\bar{q} + \kappa(j;t)\,\frac{\bar{q}}{z} + g\frac{z}{2} \tag{3}$$

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

$$z(j;t) = \left(\frac{2\bar{q}\kappa(j;t)}{g}\right)^{1/2} \tag{4}$$

Consumers buy more per trip when the travel costs are high (to economize on the number of trips) or when the storage costs are low (to take advantage of the durability of the good). Note that, conditional on $\bar{q}$, variation in g induces variation in the frequency $\bar{q}/z$. Using this expression in (3), we can rewrite the cost minimization uniquely as a function of distance. The optimal distance traveled then satisfies

$$-p'(j)\,\bar{q} = \frac{1}{2}\kappa(j;t)^{-1/2}\,\kappa_j(j;t)\,(2\bar{q}g)^{1/2} \tag{5}$$

In choosing the optimal distance, consumers intuitively balance price savings with travel and inventory costs. A marginally longer trip makes consumers save $-p'(j)$ per unit purchased;[16] however, they pay more in travel costs and inventory costs (since with longer distances they optimally buy larger batches). A higher storage cost effectively raises the marginal cost of distance in (5): as g increases, inventories are

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

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

[16] In equilibrium, the price will decrease in j, so that $-p'(j) > 0$.

more expensive, the optimal batch size in (4) is smaller and the optimal number of trips $\bar{q}/z$ grow. Quite naturally, consumers choose to travel closer and more frequently when storability/durability decreases, everything else equal.

2.3 Equilibrium price

We now solve for the equilibrium price function and the associated spatial distribution of production.[17] The price function will turn out to be the solution to a second order differential equation. We make three assumptions that allow for analytic results (and point out where they are helpful): we specialize the travel cost function to $\kappa(j;t) = (jt)^2$, we assume that $t \sim Uniform[1,2]$, and set $\beta = 1/2$ in the production function.

Our first assumption lets us write the first order condition (5) for a consumer t as,

$$t = -p'(j) \left(\frac{\bar{q}}{2g}\right)^{1/2} \tag{6}$$

This equation describes the solution to an optimal assignment problem: for a given (monotonically decreasing and convex) price function $p(j)$, it implicitly matches each consumer type t to a unique location $j(t)$, with $j'(t) < 0$.

In equilibrium, demand and supply of goods are equal in all j where consumers choose to travel. Using (6) and the distributional assumptions on t, we can write the cumulative distribution function $\Pr\left\{j < \bar{J}\right\}$ of consumers traveling up to a distance $\bar{J}$, and its density $n\left(\bar{J}\right)$, as

$$\Pr\left\{j < \bar{J}\right\} = \Pr\left\{t \geq -p'\left(\bar{J}\right)\left(\frac{\bar{q}}{2g}\right)^{1/2}\right\} = 2 + p'\left(\bar{J}\right)\left(\frac{\bar{q}}{2g}\right)^{1/2} \Longrightarrow n\left(\bar{J}\right) \equiv p''\left(\bar{J}\right)\left(\frac{\bar{q}}{2g}\right)^{1/2} \tag{7}$$

The total quantity demanded at j is then $N\bar{q} \cdot n(j)$. Equating this demand to supply (2), the equilibrium in location j requires

$$p''(j) = \alpha^2 p(j), \text{ with } \alpha \equiv \alpha_0 \frac{g^{1/4}}{\bar{q}^{3/4} N^{1/2}}, \tag{8}$$

where α_0 depends on parameters.[18] This condition must hold for any location j where the good is produced. Equation (8) is a second order differential equation in the price function $p(j)$.

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

To find the equilibrium price function, we impose that the consumers with the lowest ($t = 1$) and the highest ($t = 2$) travel costs choose the maximum distance, $j(1) = j_{\max}$, and the minimum distance, $j(2) = 0$, respectively; and that, at $j = j_{\max}$, the price of the product is zero (otherwise, rents $R(j)$ would

[17] All necessary derivations in this subsection and the next are reported in Appendix B, p. 42.

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

be positive, and some firms would have an incentive to enter slightly farther, paying zero rent). The equilibrium price function is given by

$$p(j) = \frac{1}{2^{1/2}\alpha_0} \cdot \bar{q}^{1/4} g^{1/4} N^{1/2} \cdot [\exp\{\alpha(j_{\max} - j)\} - \exp\{-\alpha(j_{\max} - j)\}] \tag{9}$$

The price decreases with distance for $j \in [0, j_{\max}]$. The value $j_{\max}$ is an implicit and decreasing function of α.

2.4 Aggregate Implications

We now explore some equilibrium implications of this stylized model.

Gravity holds. Expenditure at a given place j is given by

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

where the last equality uses (7) and (8).[19] Since the price is decreasing in j, expenditure decreases with distance and gravity holds. Analogously to continuous–types labor market models of assignment,[20] this framework delivers a non-constant distance elasticity of expenditure. This elasticity is generated by the complementarity between j and t, and the distribution of t.[21]

Gravity is stronger when storage costs are higher. Consider the slope $[X(j_{\max}) - X(0)]/j_{\max} = -X(0)/j_{\max}$. When g is higher, the marginal cost of distance for all consumers grows and the willingness to take long trips shrinks. This implies that the marginal plot of land is no longer viable for production, and $j_{\max}$ falls.[22] Since output is still fixed at $N\bar{q}$ but there is less land, more demand must be concentrated at shorter distances, in particular at $j = 0$, so that $p(0)$ grows.[23] The average slope of the expenditure is then higher when g grows.[24]

The average frequency of transactions grows when storage costs are higher. Consider the expression for the optimal batch size z in (4): as g grows, a consumer will reduce the batch size z for any distance traveled. She will also reduce the distance traveled $j(t)$, since the marginal cost of distance is higher. Hence, z unambiguously drops, and the frequency $\bar{q}/z$ unambiguously rises for each individual, and on average.[25] Conditional on $\bar{q}$, a higher frequency proxies for a higher g.

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

[20] See for example Sattinger (1979), Costinot and Vogel (2010), Monte (2011).

[21] A related setting is studied in Karádi and Koren (2017): they focus on the general equilibrium implications of sectoral location choice (our model is partial equilibrium in that we only study equilibrium in one product market), where, however, the impact of distance is modeled as a classic iceberg decay; hence, expenditure declines log-linearly with distance. From this perspective, our framework is related to demand–side nonlinear pricing models in international trade: for example, Fieler (2011) or Fajgelbaum, Grossman and Helpman (2011).

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

[23] Using (9), it's easy to see that $p(0) \propto g^{1/4}[\exp\{\alpha j_{\max}\} - \exp\{-\alpha j_{\max}\}]$, which increases with g since we show in Appendix B that $\alpha j_{\max}$ is constant.

[24] Since the equilibrium expenditure function has a varying distance elasticity over j, it is difficult to prove that the slope becomes steeper at every j. However, Lemma 2 in page 42 shows that over $j \in [0, j_{\max}]$, 1) the unweighted average slope $X'(j)$, 2) the average slope $X'(j)$ weighted by the number of agents $n(j)$, and 3) the average slope of $X'(j)$ weighted by the expenditure at j, $X(j)$, all become more negative as g grows.

[25] See Lemma 3 on p. 44 for a proof.

A crucial consequence of the last two implications is that gravity is stronger in high-frequency sectors.

In response to an increase in the number of consumers, 1) total employment grows more, and 2) the average distance of output to consumers grows less, when storage costs are higher. The total employment and the average distance of output to consumers in the sector are given by

$$L_{eq} = \int_0^{j_{\max}} L(j)\,dj = \frac{\bar{l}}{Aw^{1/2}} \cdot g^{1/4} N^{3/2} \tag{10}$$

$$D_{eq} = \frac{\int_0^{j_{\max}} jQ(j)\,dj}{Nq} \equiv \bar{d} \cdot \frac{N^{1/2}}{g^{1/4}} \tag{11}$$

where $\bar{l}$ and $\bar{d}$ depend on parameters not including g or N.[26] The intuition is simple. If population N increases, total output $N\bar{q}$ must grow in the same proportion, by the assumption of fixed $\bar{q}$. This expansion in demand – and production – requires both more workers L_{eq} and more distant land $j_{\max}(\alpha)$. The incentive of producers to economize on distant land, however, is stronger when g is higher: with high storage costs, long distances are particularly expensive for consumers; hence, employment must grow relatively more, and distance relatively less, to accommodate the increase in production.

2.5 Confounding Factors

We have so far described the simplest possible framework to illustrate our logic. In this subsection, we discuss the theoretical implications of two features of the data that might be relevant in our empirical analysis. First, we have assumed that labor supply is infinitely elastic at a given wage; however, local labor supply may be upward sloping in wages, for example, via commuting links;[27] this effect may slow down the growth in employment in response to exogenous changes. Second, while we can, in theory, consider the effect of an exogenous change in the population, in practice variation in population may be associated with changes in local productivity, which also affect employment. Hence, simple linear regressions may suffer from endogeneity bias.

Let us assume that the labor supply has a finite elasticity $\varepsilon > 0$ with respect to wages w, i.e., $L^s = Nw^{\varepsilon}$. Equation (10) now just describes the labor demanded by firms L^d, rather than equilibrium employment. Let us also assume that the equilibrium population and productivity are related via $N = \bar{N}A^{\zeta}$, where $\bar{N}$ is an exogenous component, and $\zeta \geq 0$ captures in a reduced-form way the positive relationship between local productivity and population. This relation may come via agglomeration externalities or because high productivity reduces the local consumer price index and attracts more residents. Equating labor demand L^d to labor supply,

$$L^d = \bar{L} \cdot g^{1/4} \bar{N}^{3/2} A^{3\zeta/2-1} w^{-1/2} = \bar{N}A^{\zeta}w^{\varepsilon} \tag{12}$$

[26] See Lemmas 4 and 5 and on pp. 45 and following for a proof.

[27] See for example Monte, Redding and Rossi-Hansberg (2018).

This equation determines an equilibrium wage,[28] which we can then substitute in eq. (10) to obtain

$$L_{eq} = \hat{L} \cdot g^{\frac{1}{2}\frac{\varepsilon}{(2\varepsilon+1)}} \bar{N}^{\frac{3\varepsilon+1}{2\varepsilon+1}} A^{\frac{(3\varepsilon+1)\zeta-2\varepsilon}{2\varepsilon+1}} \tag{13}$$

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

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

In other words, a simple comparison of low to high population (or population density) places may also be an implicit comparison of low to high productivity places; and a higher productivity, everything else equal, implies employment that grows relatively less, or falls relatively more, in high g sectors than in low g sectors. This direction would be the opposite indicated by our simple model.

Note, however, that if we were able to induce exogenous variation in $\bar{N}$, we should still expect a positive cross partial between population and g, since:

$$\frac{\partial L_{eq}}{\partial g \partial \bar{N}} > 0 \tag{14}$$

This cross-partial derivative is positive even in the presence of labor market equilibrium feedbacks (i.e., our intuition carries through to a more general data-generating process where equilibrium employment responds to local labor supply and wages).

Armed with these intuitions, we turn to an empirical exploration of consumer mobility and its consequences on local outcomes.

3 Consumer Mobility

In this section, we present basic facts about consumer mobility. We argue that consumer mobility is low and that distance appears a primary friction. Expenditure declines quickly with distance. This decline is heterogeneous across sectors, is mostly explained by extensive margins of adjustment, and is related to the average frequency of transactions within a sector.

3.1 Data Description

We describe consumers' geographical mobility patterns using a large proprietary dataset containing a sample of credit card transactions from a major financial institution. These transactions occurred roughly between March and October 2003. A transaction record contains, among other things, an exact date, an account ID, the amount spent, and a Merchant Category Code (MCC), which we will refer to as a "broad category," in this classification's terminology, or simply "sector". A transaction record also

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

contains information (to be processed) on the purchase location. This location can only be identified at the level of Census incorporated place or county subdivision. In addition to all distinct transactions, we have information on the account itself, including self-reported income, age, and the associated ZIP code.[29] We aggregate residence ZIP codes to the level of the purchase location using a correspondence from the GeoCorr database from the Missouri Census Data Center. After cleaning the data, we have 1,722,873 transactions for 71,377 accounts (see Appendix A for a complete description of data cleaning and processing). The average transaction is 68 dollars, and total purchases amount to around $116 million. Table 1 gives a breakdown by 21 broad categories. The largest categories in terms of observations are Gasoline Services, Food Stores, Miscellaneous Retail, and Eating and Drinking Places.[30]

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

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

3.2 Consumers Visit Few Locations

We start our exploration by considering how far consumers travel across locations for purchases. In the raw data with all transactions, we count expenditure flows between around 18 thousand unique locations. Since we want to describe "day-to-day" mobility and purchasing behavior, we focus, in much of the analysis, on transactions that occur within a distance of 120 km between the location of residence and

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

[30] Table C.1 in the Appendix (page 48) shows summary statistics by state of purchase.

sale.[31] Among all the location pairs within this threshold, we observe around 1.5 million transactions between only around 121 thousand pairs (2.8% of the possible pairs). There are 12.3 transactions per pair on average, and two transactions for the median pair. Overall, the matrix of residence-sales location purchases appears sparse.

To dig deeper into this apparently low mobility, we construct a residence location–level dataset. For each residence, we compute 1) the total number of locations visited by all consumers living in that residence; 2) the total number of locations in the data within 120 km; 3) the average distance from the residence location to these locations; and 4) the share of locations visited out of those available. The four rows of Table 2 report summary statistics on these items. The first row shows that consumers in the median residence visit only 7 distinct sales locations during the sample period. The second and third rows indicate that these residences vary both in terms of how many locations are close-by and in terms of their average distance to them. The fourth row shows that, regardless of the source (few opportunities or long distances), the overall fraction of available locations where purchases occur is also small.

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

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

In Table 3 , we then ask what accounts for this low mobility. In column 1 of Table 3 we regress the log of the number of sales locations visited on the log of the number of sales locations available and find an elasticity of 0.55: overall, the number of visited locations grows at about half the pace of the available ones. Distance has a strong role: controlling for the number of available sales locations, a 1% increase in the average distance to those locations is accompanied by a 2.4-2.6% decrease in the number of locations visited (columns 2 and 3).[32] The next section characterizes distances traveled more in detail.

3.3 The Distance Traveled Varies by Sector

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

The median transaction in the data occurs about 9 km from home. There is a large dispersion around this typical value: the first 25% of transactions occur within the same location, while the third quartile is around 30 km. A long right tail of high distances is likely due to account holders traveling outside town

[31] Distance is always computed between the centroids of two locations using the Haversine formula. We identify and exclude on-line transactions, where the distance is not meaningful. Monte, Redding, and Rossi-Hansberg (2018) find this threshold to be one where gravity in home-to-work commuting flows has a change in slope, so it is a natural cutoff.

[32] In Tables C.2 and C.3 of Appendix C.2 (page 47), we repeat this analysis using a sample of users with at least one transaction every two days. The fraction of locations visited has a similar distribution. In a regression analysis, distance has twice the impact of the number of available locations on the total number of locations visited.

Table 3: Locations available and locations visited

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

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

for work or tourism.[33] Figure 1 shows select percentiles of the distances at which transactions occur, by sector. The heterogeneity in the distance traveled is very significant: moving from the sector at the 10th percentile of the median distance traveled to the sector at the 90th, such median distance 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.[34]

The distance traveled by consumers is a combination of their willingness to travel (as mediated by their optimal shopping behavior) and supply conditions like the density of producers. We return to this consideration in robustness exercises at the end of this section. Having characterized distances traveled, we now analyze how they relate to expenditure flows.

3.4 Gravity in Consumer Expenditure

A substantial literature has documented the decay of goods' trade flows with distance at international and intra-national levels; these studies have limited implications for the behavior of consumers, who buy merchandise from producers only infrequently and consume services in addition to goods.[35] We address this gap in two steps: first, we document that gravity also holds for consumers' behavior; second, we analyze the margins of this decline across individuals and point out a stylized feature of the data consistent with our simple model: gravity is stronger where the frequency of transactions is higher.

[33] We report more statistics on the untruncated data in the Appendix. Tables C.4 and C.5, in pages 50 and 51 respectively, show percentiles in the distribution of transaction distances by sector, unweighted and weighted by value of the transaction. People spend the typical dollar farther than where the typical transaction occurs, as reflected in rightward shifts in the value-weighted distributions.

[34] 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: consistently, we find for the same industry that 75% of the transactions occur within around 11 miles.

[35] See, e.g., Hillberry and Hummels (2007) or Disdier and Head (2008). Intranational flows of goods typically record only firm-to-firm transactions.

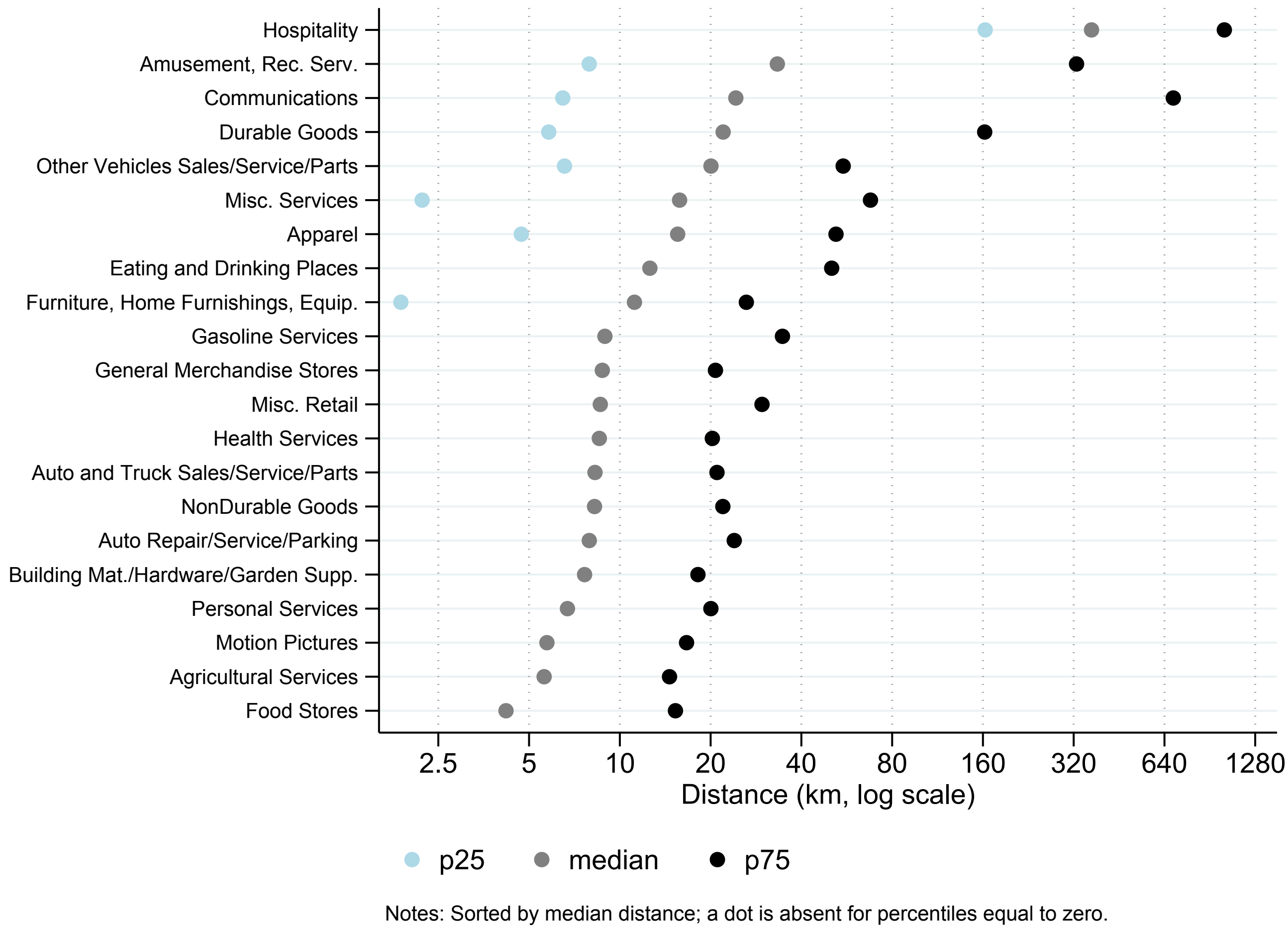

Figure 1: Distances traveled by sector (select percentiles)

3.4.1 Expenditure patterns display gravity

We start our exploration of gravity in consumer behavior by investigating how quickly total expenditure decays with distance.[36] For a given sector, we denote with X_{hs} the total expenditure of consumers residing in location h on merchants selling in location s. Except where noted otherwise, we restrict from now on the analysis to transactions occurring at distances up to 120 km. We make full use of the information available in the data comparing 1) expenditure inside vs. outside one's place of residence, and 2) the decline in expenditure across merchants at different distances from the residence location. First, we estimate the change in expenditure associated with shopping out of the home residence ("out-of-home"):

$$\log X_{hs} = \alpha + \phi^{(h)} + \phi^{(s)} + \eta \times \mathbf{1}_{(h \neq s)} + \varepsilon_{hs} \tag{15}$$

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

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

$$\log X_{hs} = \alpha + \phi^{(h)} + \phi^{(s)} + \delta \log dist_{hs} + \varepsilon_{hs} \tag{16}$$

In this equation, $dist_{hs}$ is the distance between the centroids of h and s; this regression includes only pairs where $h \neq s$ because we do not measure the distance of transactions within the residence location. In both equations, α is a constant, and a set of origin and destination fixed effects, $\phi^{(h)}$ and $\phi^{(s)}$, controls for unobserved location-specific differences in factors such as size, transportation infrastructure, productivity, or intensity of competition (Anderson and Van Wincoop, 2003). These two approaches highlight complementary features of the data. The coefficient η in Equation (15) measures the expenditure drop associated with visiting the average out-of-home location; hence, it shows the importance of very short trips, for which distance is poorly measured. The coefficient δ in Equation (16) shows the elasticity of expenditure to distance considering only out–of–home expenditure flows, where instead distance can be measured.

We estimate Equations (15) and (16) across all sectors. We find, unsurprisingly, very clear distance effects. Estimating (15), the expenditure in the average location out-of-home is only about 8.8% of the average expenditure at home ($\eta = -2.435$, robust s.e. 0.021).[37] When we estimate (16), we find a slope of -1.051 (robust s.e. of 0.006), in line with estimates in the trade literature.[38] A comparison of these two coefficients shows that a large decay already occurs at very short distances.

These pooled estimates mask large differences across sectors. Table 4 shows the coefficients of η (column 1) and δ (column 4) when we estimate Equations (15) and (16) by sector.[39] Sectors in this table are ordered by the out-of-home dummy in column 1 (this ordering will be kept throughout the paper for ease of reference). The strong decay at short distances is pervasive across sectors. However, the decay is heterogeneous: in sectors like Food Stores, the expenditure in the average location out-of-home is around 9% of the expenditure at home; this fraction grows to 20% for Eating and Drinking Places, to 38% for Personal Services, and 91% for Durable Goods.

The impact of distance as measured by estimates of (16) is consistent with this picture: the correlation between the two sets of coefficients across sectors is 0.68. However, the much smaller distance coefficients in these estimates are notable: for the typical sector, a 1% increase in distance is associated with a 0.41% decrease in expenditure, and almost all coefficients are below the benchmark value of approximately 1 for international trade. We conclude that most of the decline in expenditure happens at short distances as measured in (15), which we will focus on for much of the remaining analysis.

Column (7) in Table 4 reports, for each sector, the simple average of the number of purchases per account in the transaction data. Our simple model above suggests that the frequency of transactions may serve an inverse proxy for the storability/durability of a sector's output, g: a lower storability/durability

[37] Using all data rather than only transactions within 120 km, we find $\eta = -2.545$ (robust s.e. 0.0223).

[38] This slope is not particularly sensitive to changes in the cutoff. See Appendix C.4, page 52, for further discussion.

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

increases inventory costs and induces shorter and more frequent trips. The model also suggests that the observed frequency of transactions can be influenced by the overall importance of a sector in consumers' expenditure, $\bar{q}$. To capture this intuition, we develop a correspondence between the MCC classification and the items in the Consumer Expenditure Survey from the Bureau of Labor Statistics. We report the budget shares in this data (CEX expenditure shares in what follows) among our sectors in 2003 in Column (8). We note for now the tendency of sectors with stronger gravity to be purchased more frequently. We return to this aspect in more detail after we have a better understanding of which margins account for the expenditure drop over space.

Table 4: **Decline in expenditure**

	Out of Home			Gravity			Frequency of	CEX
Category	coeff	pv	obs.	coeff	pv	obs.	transactions	share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Food Stores	-2.23	0.00	22,649	-0.85	0.00	18,632	7.53	0.14
Gasoline Services	-2.08	0.00	39,666	-0.60	0.00	34,615	8.86	0.06
General Merchandise Stores	-1.78	0.00	26,837	-0.93	0.00	23,932	5.14	0.02
Misc. Retail	-1.70	0.00	34,052	-0.65	0.00	30,042	5.25	0.06
Eating and Drinking Places	-1.57	0.00	34,504	-0.56	0.00	31,022	5.93	0.10
Building Mat./Hardware/Garden Supp.	-1.40	0.00	14,185	-0.73	0.00	11,604	4.15	0.04
Auto Repair/Service/Parking	-1.25	0.00	4,414	-0.40	0.00	3,013	1.83	0.03
NonDurable Goods	-1.16	0.00	978	-0.65	0.00	758	1.68	0.01
Health Services	-1.12	0.00	5,134	-0.33	0.00	3,910	2.16	0.05
Apparel	-1.10	0.00	15,918	-0.53	0.00	14,066	2.91	0.07
Furniture, Home Furnishings, Equip.	-1.07	0.00	12,286	-0.57	0.00	10,734	2.33	0.07
Auto and Truck Sales/Service/Parts	-1.04	0.00	7,298	-0.33	0.00	5,508	1.98	0.16
Motion Pictures	-1.04	0.00	1,922	-0.34	0.00	1,248	2.16	0.02
Amusement, Rec. Serv.	-1.03	0.00	2,958	-0.23	0.00	2,329	2.03	0.04
Personal Services	-0.96	0.00	5,203	-0.31	0.00	3,760	2.46	0.01
Misc. Services	-0.92	0.06	220	0.91	0.02	116	1.57	0.03
Communications	-0.89	0.00	424	-0.41	0.01	263	1.36	0.04
Agricultural Services	-0.88	0.00	552	0.42	0.11	190	1.86	0.02
Other Vehicles Sales/Service/Parts	-0.68	0.41	257	-0.59	0.08	128	1.64	0.00
Hospitality	-0.64	0.01	1,392	-0.14	0.08	1,158	1.53	0.02
Durable Goods	-0.09	0.90	79	1.11	0.67	15	1.64	0.03

3.4.2 Gravity and frequency

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

$$X_{hs} = \underbrace{N_{hs}}_{\text{account margin}} \times \underbrace{\bar{x}_{hs}}_{\text{expenditure margin}} = \tag{17}$$

$$= \underbrace{N_{hs}}_{\text{account margin}} \times \underbrace{f_{hs}}_{\text{frequency margin}} \times \underbrace{\bar{x}_{hs}/f_{hs}}_{\text{batch size margin}} \tag{18}$$

Equation (17) 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, (18) suggests that lower expenditure per account on average can arise either because each transaction is smaller (the "batch size" margin) or because consumers transact less often (the "frequency" margin). When we re-estimate Equation (15) with the left side being the log of each of these three terms, the coefficients on the out-of-home dummy add up to the overall coefficient η reported in column 1 of Table 4 (and similarly for Equation (16)).[40]

Figure 2 shows the results of this decomposition for (15). The length of each bar corresponds to column 1 in Table 4. We find two broad messages.

First, most of the decline in expenditure over space is due to fewer people traveling outside, or people taking less frequent trips. The blue bar measures the contribution of the "account" margin. For the typical sector, 72% of the drop in out-of-home expenditure is associated with fewer people traveling outside, rather than to people spending less on average for out-of-home transactions.[41]

The remaining part of each bar measures the decline due to lower average expenditure per account. The gray section indicates that the average expenditure per account drops outside of home almost exclusively because of the "frequency" margin: consumers spend less on average out-of-home because they choose to travel outside less frequently, not because they spend less per transaction. The drop in the average transaction value (the "batch size" margin) has a limited role in most cases. Tables C.8 and C.9 in the Appendix (p. 54) show that the combination of the "account" and "frequency" margins typically contributes 90%-95% of the decline in expenditure. As a benchmark, Hillberry and Hummels (2007) find that, for firm-to-firm shipments within U.S., for short distances the extensive margin explains almost the totality of the decay.

Second, Figure 2 suggests that a large part of the heterogeneity in gravity seems associated with heterogeneity in the frequency margin and not associated with the "batch size" margin – i.e., the length of the bar varies because of variation in the gray section.[42] This fact 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

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

[41] Tables C.6 and C.7, in the Appendix (p. 53), show the actual values of the "account" and "expenditure" margins with associated p-values for both Equations (15) and (16).

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

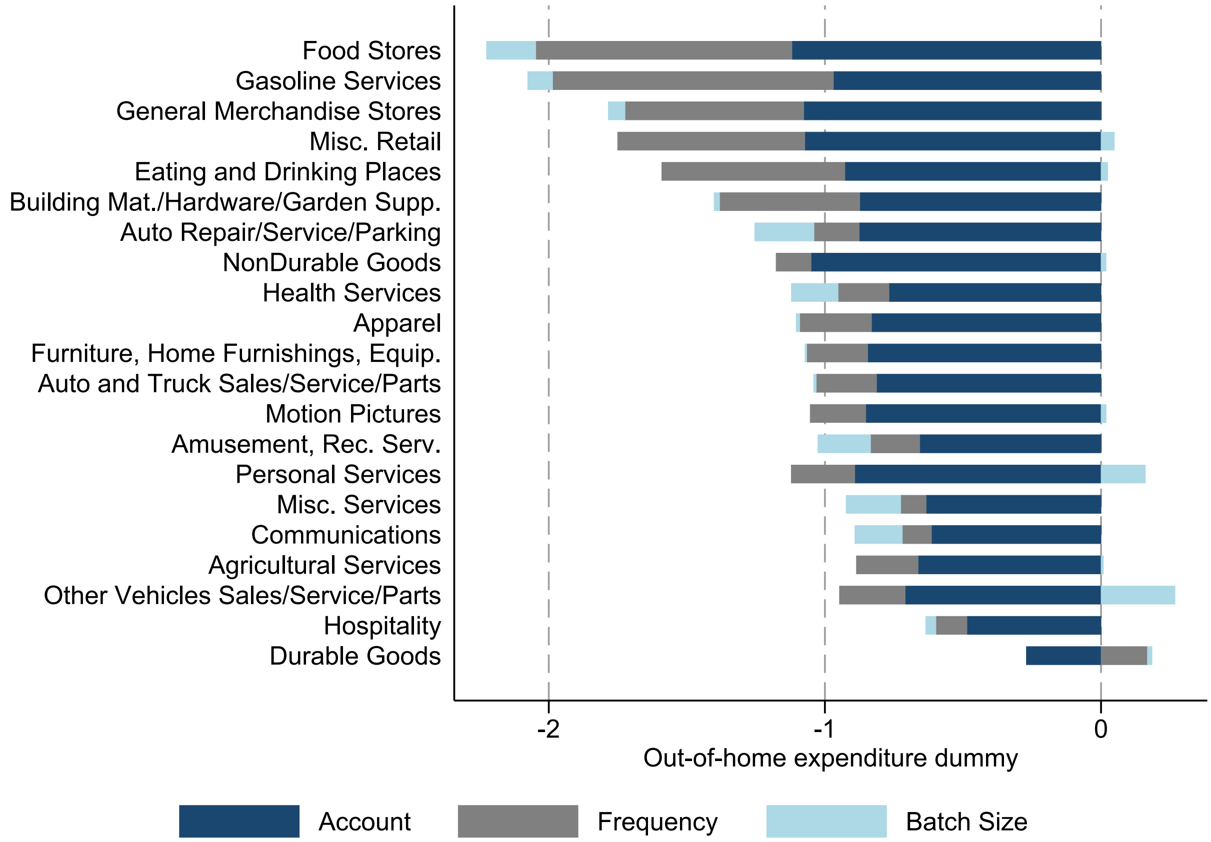

Figure 2: Margins in the out-of-home expenditure drop

for the sectors where the out-of-home dummy is significantly different from zero at 10% level. A simple regression line through this data has a slope of -.69 (robust s.e. 0.07) and an R^2 of 0.86.[43] 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.

In line with an implication of our simple model, expenditure decays faster in more frequently transacted sectors. Our interpretation of this relation is that when storage costs are high, the average inventory held shrinks: this reduction is achieved by purchasing smaller batches, but more frequently. Since travel is expensive, however, a higher frequency can only be optimal with reduced travel distances. Hence, across sectors, if storage costs are higher, the frequency of purchase should grow, and the expenditure should decline faster with space.

Our interpretation of this figure relies on two orders of considerations. First, we are using the frequency of transactions as a proxy for storability/durability. A higher frequency however, might also arise from

[43] The figure excludes Durable Goods and Other Vehicles Sales/Service/Parts, an outlier. Using all estimates, the regression line would have a slope of -0.77 (robust s.e. 0.09), with $R^2 = 0.78$. Figures C.2 and C.3 in the Appendix, starting at page 55, replicate this figure for all the sectors and for the impact of distance using eq. (16). We have also experimented with an alternative measure of frequency that gives more weight to users which spends more overall in the data, with very similar results.

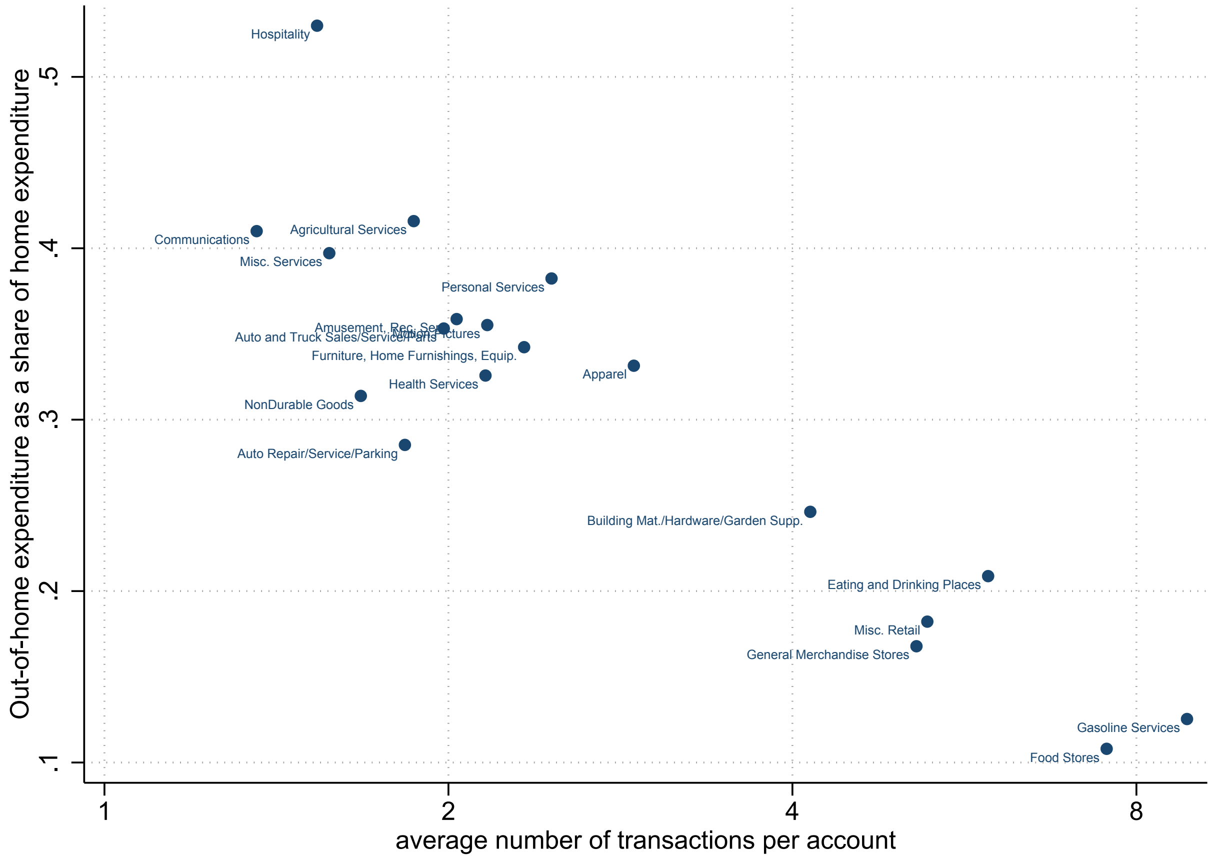

Figure 3: Drop in expenditure out of home

a larger expenditure share. We find that the linear pattern in Figure 3 does not appear to be driven by this alternative explanation: after controlling for the CEX expenditure share, a linear fit for the figure has a very similar slope of -0.68 (robust s.e. 0.06), and leaves the R^2 basically unaffected; the expenditure share has a coefficient of -.33 (robust s.e. of 0.9). Hence, the variation in the frequency of transactions associated to gravity does not appear to be induced by heterogeneous budget shares.

Second, our interpretation of Figure 3 implicitly assumes that consumers actually make travel distance decisions, and that these decisions are in part related to the storability/durability of a sector's output. Although our intuition about consumer's travel likely conforms to everyday experience, some supply-side characteristics of different industries, like fixed costs, may make some sectors' establishments denser in space, and render frequent trips less expensive. This scenario might produce patterns consistent with those in Figure 3. In two exercises in Appendix D, we directly examine the individual-level spatial distribution of transactions during the eight months of our data. This short span allows us to assume that the retail environment (location and size of establishments) is stable. We find evidence consistent with explicit consumer decisions in travel distance that are related to the average frequency of transactions across sectors, after controlling for CEX expenditure shares.

We first examine the role of individual agents' characteristics, like age and income, on the out-of-home fraction of trips across sectors. After controlling for sector and ZIP code or individual fixed effects, as well as a number of other factors, we find that an individual with higher income makes at-home purchases relatively more in high-frequency sectors, compared to an observationally equivalent, lower-income individual. Econometrically, the coefficient on the simple interaction between income and frequency of transactions is negative and significant.

We also compare the behavior of the same individual across sectors under two travel costs regimes as a function of that individual's characteristics. In particular, we leverage daily precipitation data provided by the National Oceanic and Atmospheric Administration (Menne et al., 2012). For each centroid of a residence location in our data, we identify the closest weather station among the roughly twelve thousand disseminated across the United States, and we reconstruct local weather conditions at the time of each purchase. We then compare the fraction of out-of-home transactions of the same individual during rainy vs. non-rainy days across sectors, and relate this difference to individual characteristics. We allow here as well for a rich set of controls. We find that an individual with higher income responds to the same "rain" shock by making purchases in high-frequency sectors relatively more local, as compared to an observationally equivalent, lower-income individual. Econometrically, the coefficient on the triple interaction between income, frequency of transactions, and a "rain" dummy is negative and significant.

In both cases, we find that agents' income modulates the spatial distribution of purchases in a way that is related to the sector average frequency of transactions. In these exercises, the retail environment is plausibly stable and confounding factors are accounted for with a detailed set of controls; we conclude that the differential impact of individual income on observed travel across sectors is likely to come from an active consumer decision.

These results do not speak to the equilibrium impact of consumer mobility on the composition of local employment and establishment density. We turn to this question next.

4 The Local Structure of Consumption Industries

To quantify the influence of consumer mobility on the local structure of consumption industries, we leverage an intuition we have provided in Section 2.4. In response to a larger local population, we might expect to see local employment in high storage-cost sectors grow relatively faster than employment in sectors with low storage costs: firms economize on consumer travel time by limiting the amount of distant land used, substituting land with labor to a greater extent the lower is the storability/durability of the sector. In addition, if savings in travel time are at the root of this behavior, we might expect to see employment growth driven by a higher density of stores (that is, a reduced average distance between consumers and stores).

In this section, we consider the empirical counterpart of that relation. We study the impact of differences in population on county–sector outcomes as a function of the sector's average number of transactions, our simple proxy for storage costs. We see the forces we describe playing out in a long-run equilibrium, after the entry-exit margin of new establishments has been allowed to adjust. Hence, our main empirical examination will leverage cross-sectional differences across space and sectors. In particular, we estimate regressions of the form

$$\ln y_{sct} = \alpha + \beta \ln pop_{ct} + \gamma \ln freq_s \times \ln pop_{ct} + \beta' X + FE + \varepsilon_{sct} \tag{19}$$

In this regression, s indexes MCC sectors, c indexes counties, and t denotes calendar years ($t = 2007$ and 1998). The dependent variable $\ln y_{sct}$ may assume three values. We first use log employment in s, c, t, $\ln emp_{sct}$: to construct it, we have started with data in the relevant years from County Business Patterns, and have developed a correspondence between NAICS 6 digits and MCC codes. Always using County Business Pattern data, we also explore the response of the number of local establishments $\ln y_{sct} = \ln n_{sct}$, and employees per establishment $\ln y_{sct} = \ln (emp_{sct}/n_{sct})$; $\ln pop_{ct}$ is the county log population in the year from the County Economic Profile of the Bureau of Economic Analysis, and $freq_s$ is the average frequency of transactions by sector in the credit card data, as reported in column (7) of Table 4 above. The vector X collects other controls like the county's average personal income per capita, from the County Economic Profile of the Bureau of Economic Analysis, the county land's area, and may include other controls as robustness, as described below. Importantly, we always include the interaction of log population with the CEX expenditure share: hence, we interpret the frequency of transactions as a proxy for storability/durability. Finally, FE is a set of fixed effects, varying across specifications (sector fixed effects are always included); and ε_{sct} is a stochastic unobserved term.

4.1 Cross-Sectional Correlations

We start by estimating eq. (19) with OLS for 1998 and 2007. Table 5 reports the results. In the first two columns, the dependent variable is county-sector log employment; columns (3) and (4) use total number of establishments, and columns (5) and (6) use the average number of employees per establishment. Each group of two columns differs by the set of fixed effects used: columns (1), (3) and (5) use sector fixed effects and year fixed effects; columns (2), (4) and (6) allow for commuting zone-year and sector-year dummies, i.e., control for heterogeneous time trends within sectors and within commuting zones.

The regression results show that population enters positively. Moreover, sectoral employment grows less with population in sectors that are transacted more frequently. Columns (3)-(6) indicate that this slower employment growth occurs about equally via fewer establishments and a lower number of employees per establishment.

This cross-sectional correlation runs contrary to our intuition, which would suggest a positive interaction between population and frequency. The analysis in Section 2.5 indicated that unobserved factors like local productivity might generate a negative interaction between population and frequency in equilibrium. We then inspect the source of the negative interaction sign in Table 5. In particular, we consider the behavior of counties with different initial characteristics that may be related to productivity. We group counties in 5 bins according to their initial population density in 1998: if county population and density are related to productivity (e.g., Combes et al., 2012), we might expect this coefficient to vary with the set of counties included in the regression. We then re-estimate columns (2), (4) and (6) in Table 5 progressively for the set of counties up to and including the q^{th} quintile, with $q = 1, ..., 5$. The left panel

Table 5: **Local outcomes and frequency of purchase (OLS)**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log land area	0.125***	-0.023	0.122***	0.011	0.003	-0.034***
	(0.010)	(0.016)	(0.007)	(0.011)	(0.006)	(0.009)
Log income per capita	1.113***	1.006***	0.935***	0.813***	0.178***	0.193***
	(0.035)	(0.044)	(0.025)	(0.030)	(0.020)	(0.025)
Log population	1.209***	1.386***	0.907***	1.015***	0.303***	0.371***
	(0.007)	(0.010)	(0.005)	(0.006)	(0.004)	(0.006)
Log population × log frequency	-0.077***	-0.079***	-0.041***	-0.041***	-0.036***	-0.038***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Log population × CEX expenditure share	0.388***	0.395***	0.623***	0.627***	-0.235***	-0.231***
	(0.050)	(0.050)	(0.028)	(0.028)	(0.041)	(0.041)
Constant	-13.469***	-11.780***	-12.185***	-10.544***	-1.284***	-1.236***
	(0.242)	(0.352)	(0.159)	(0.242)	(0.141)	(0.198)
Sector Fixed Effects	Yes	No	Yes	No	Yes	No
Year Fixed Effects	Yes	No	Yes	No	Yes	No
Sector-Year Fixed Effects	No	Yes	No	Yes	No	Yes
Commuting Zone-Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-square	0.84	0.86	0.89	0.91	0.52	0.54
N	121,336	121,336	121,336	121,336	121,336	121,336

Standard errors clustered at county level in parentheses.

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

of Figure 4 plots the coefficient γ_q on the interaction between population and frequency including progressively more quintiles. The panel shows that in low-density counties, a marginal increase in population is associated with larger employment in high-frequency compared to low-frequency sectors. This larger employment occurs entirely via an increase in the number of establishments, i.e., via a reduction in the distance of consumers to establishments within the county. Hence, we find support for the implications of our stylized model on the cross-sectoral behavior of employment in this subset of counties. As we include progressively denser (and arguably more productive) counties, however, the effect becomes smaller and then it reverses.

Our analysis of potential sources of endogeneity in Section 2.5 suggests a potential alternative strategy to examine the implications of our simple model. In particular, eq. (13) indicates that if we were able to induce variation in the exogenous component of the population $\bar{N}$, we should expect a positive cross partial between population and g. In the next subsection, we use an instrumental variable approach to approximate this variation.

4.2 Instrumental Variable Estimates

To induce variation in $\bar{N}$, we exploit the underlying geological composition of a county, namely the presence of aquifers, borrowing an intuition developed in Burchfield, Overman, Puga, and Turner (2006) and Duranton and Turner (2017). An aquifer is an underground layer of water-bearing rock, and the

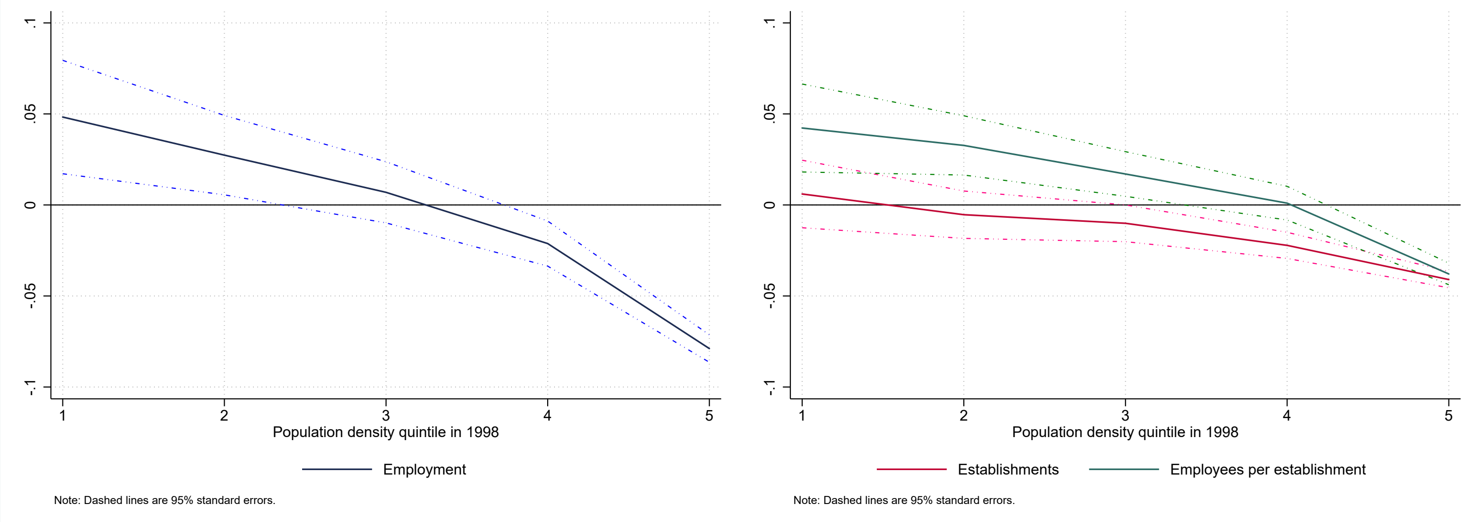

Figure 4: Slope of interaction of population and frequency (OLS)

presence of different types of aquifers induces quasi-random variation in the population residing over the territory of a county. We compute the fraction of land of a county laying over consolidated and semiconsolidated aquifers.[44] We instrument county population with those fractions, and the interaction of population and frequency of transactions with the interaction of the same percentages with the sector-level frequency of transactions. All our specifications control for the interaction of the CEX expenditure share with population, which we instrument with the product of those percentages with the CEX share. Overall, we instrument 3 endogenous variables with 6 instruments. Since the instruments are time-invariant, they induce quasi-random variation in population across counties over space but not within counties over time: hence, our estimation approach is consistent with our intention of leveraging cross-sectional differences in a long-run equilibrium.

Our main results are reported in Tables 6-8, which estimate eq. (19) using two-stage least squares. In these tables, standard errors are clustered at the county level to allow for correlation in the outcome variable, within counties, across sectors and time.[45] The strategy results in a good first stage across all specifications. Across all tables, we report the Sanderson-Windmeijer (2015) partial F-statistic for the strength of the first stage identification.

Column (1) of Table 6 shows that, after controlling for endogeneity, the sign on the interaction coefficient is of the expected sign: in 2007, when the population of a county is larger because of underlying geological reasons, the effect on employment is larger in a sector with high storage costs than in a sector with low storage costs. Moving from the minimum to the maximum average frequency changes the growth in employment by 2.2 percentage points for a 10% increase in population. This is equivalent to

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

[45] We have replicated these results with an alternative clustering: in each year, we have split counties in twenty quantiles of population, and generated groups for combinations of population class times sector. The patterns of significance in our regressor of interest are unchanged.

$100 \cdot 2.5/7.79 = 28\%$ of the estimated baseline increase in employment for the lowest frequency sector.[46] Column (2) and (3) show the same regression run in 1998 and for the stacked sample of two cross-sections with year fixed effects. Obviously, time trends may be operating differentially for different areas and sectors, and this may affect our estimates in the stacked regression. In column (4) and (5), we allow for heterogeneous time trends across sectors (both columns), and across U.S. states (column (4)) or commuting zones (column (5)). As we include more detailed geography fixed effects, the cross-sectional variation that is left for the instruments to exploit becomes narrower. Nonetheless, the coefficient on the interaction term stays positive and significant. In the most restrictive specification, where we only exploit geological differences between counties within the same commuting zone, moving from the smallest to the largest frequency sector changes the employment response by 1.39 percentage points per 10% increase in population or 11% of the baseline impact of population. In comparison, moving from the lowest to the highest CEX expenditure share sector changes the employment response by about 1.23 percentage points per 10% increase in employment or 10% of the baseline impact for the sector with the lowest share.

Table 6: **Local employment and frequency of purchase (2SLS)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log employment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.757***	0.887***	0.817***	0.803***	1.210***
	(0.097)	(0.075)	(0.083)	(0.125)	(0.084)
Log population × log frequency	0.118**	0.073*	0.096**	0.088**	0.074**
	(0.048)	(0.041)	(0.042)	(0.041)	(0.038)
Log population × CEX expenditure share	0.804	0.983*	0.903**	0.881**	0.785*
	(0.500)	(0.507)	(0.445)	(0.445)	(0.431)
Log income per capita	1.615***	1.603***	1.623***	1.769***	1.017***
	(0.140)	(0.136)	(0.134)	(0.295)	(0.202)
Log land area	0.104***	0.136***	0.120***	0.184***	-0.024
	(0.012)	(0.011)	(0.011)	(0.034)	(0.043)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.77	0.79	0.78	0.73	0.69
N	60,413	60,923	121,336	121,336	121,336
S.W. F stat: Log Population	28.03	43.35	37.78	11.92	20.14
S.W. F stat: Log Population × log frequency	20.44	24.9	25.74	13.43	16.28
S.W. F stat: Log Population × CEX expenditure share	12.81	15.89	15.29	16.55	14.88

Standard errors clustered at county level in parentheses.

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

The regressions shown so far indicate that if storage costs are reasonably proxied by the observed

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

frequency of transactions, consumers' mobility choices have economically relevant consequences on employment: in response to larger population, all sectors want their output to grow; however, the desire to economize on distance is stronger (and hence the substitution of land with employment growth is greater) in sectors where storage costs are high. To investigate how this equilibrium impact comes about, we ask next how employment is increasing in the county. In practice, an increase in local sectoral employment may be generated entirely at the intensive margin, i.e., via more employees per store. If time savings are important, however, we might expect that demand in high storage cost sectors is served by a higher density of stores, i.e., via a lower average distance between consumers and stores. Our model does not have a specific prediction about this mechanism, but suggests that the average distance between consumers and output should be smaller (and grow less with population) in high$-g$ sectors (Equation (11)); in other words, the county density of stores, an inverse proxy of the distance between consumers and establishments, should be growing faster in high g than in low g sectors.

Table 7 and 8 show that indeed, the geographical concentration of stores grows relatively faster with population in high storage-cost relative to low storage-cost sectors. In particular, Table 7 replicates Table 6 but uses the log number of establishments in a given county-sector-time as a dependent variable. Estimates on the interaction coefficient are now two to three percentage points larger and strongly significant. These results are consistent with a situation where, in response to a common increase in population, the increase in demand is more geographically concentrated for high storage costs goods and services, where people desire frequent transactions and shorter trips; the supply side then responds by increasing employment via a relatively denser presence of stores. The most conservative estimates imply that the highest storage-cost sector has 1.99 percentage points larger number of establishments relative to the lowest storage-cost sectors, when population increases 10% (about 24% of the baseline impact of population of the lowest frequency sector). In comparison, moving from the sector with the lowest to the highest CEX expenditure share implies a 2.02 percentage point increase for a 10% increase in employment, about 25% of the baseline impact of the lowest CEX share sector. Table 8 considers the reaction of establishment size. The results show that, if anything, establishments become slightly relatively smaller on average; the effect is always insignificantly different from zero.

To understand more about how the instrument works and how it flips the sign of the interaction, we re-estimate our IV regressions starting from the quintile of counties with lowest density and then progressively including all the others (mimicking the construction of Figure 4 above). The results are reported in Figure 5. The instrument operates by raising the estimated net effect of population times frequency throughout the range of densities. The sign of the interaction coefficient using the full sample (at quintile 5 in the left panel of Figure 5) switches from negative to positive because the instrumental variable approach estimates a stronger positive effect on the extensive margin of the number of stores, and a weaker negative effect on the store size.[47]

Taken together, these results paint a picture consistent with the importance of consumer mobility for

[47] A natural concern about the change in the sign between OLS and 2SLS is the presence of weak instruments. Limited Information Maximum Likelihood (LIML) estimates have better small sample properties in presence of weak instruments, and large differences between 2SLS and LIML estimates may point to instrument weakness. We replicate in Appendix E.1 the tables of Section 4.2 and show that coefficients and standard errors are indeed very similar.

Table 7: **Number of establishments and frequency of purchase (2SLS)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.471***	0.572***	0.519***	0.533***	0.789***
	(0.081)	(0.061)	(0.069)	(0.095)	(0.060)
Log population × log frequency	0.149***	0.097***	0.124***	0.114***	0.106***
	(0.038)	(0.031)	(0.033)	(0.032)	(0.029)
Log population × CEX expenditure share	1.048***	1.783***	1.419***	1.370***	1.289***
	(0.292)	(0.342)	(0.276)	(0.270)	(0.260)
Log income per capita	1.441***	1.376***	1.420***	1.439***	0.941***
	(0.117)	(0.101)	(0.106)	(0.216)	(0.133)
Log land area	0.100***	0.135***	0.118***	0.160***	0.035
	(0.011)	(0.008)	(0.009)	(0.026)	(0.028)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.81	0.84	0.82	0.78	0.76
N	60,413	60,923	121,336	121,336	121,336
S.W. F stat: Log Population	28.03	43.35	37.78	11.92	20.14
S.W. F stat: Log Population × log frequency	20.44	24.9	25.74	13.43	16.28
S.W. F stat: Log Population × CEX expenditure share	12.81	15.89	15.29	16.55	14.88

Standard errors clustered at county level in parentheses.

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

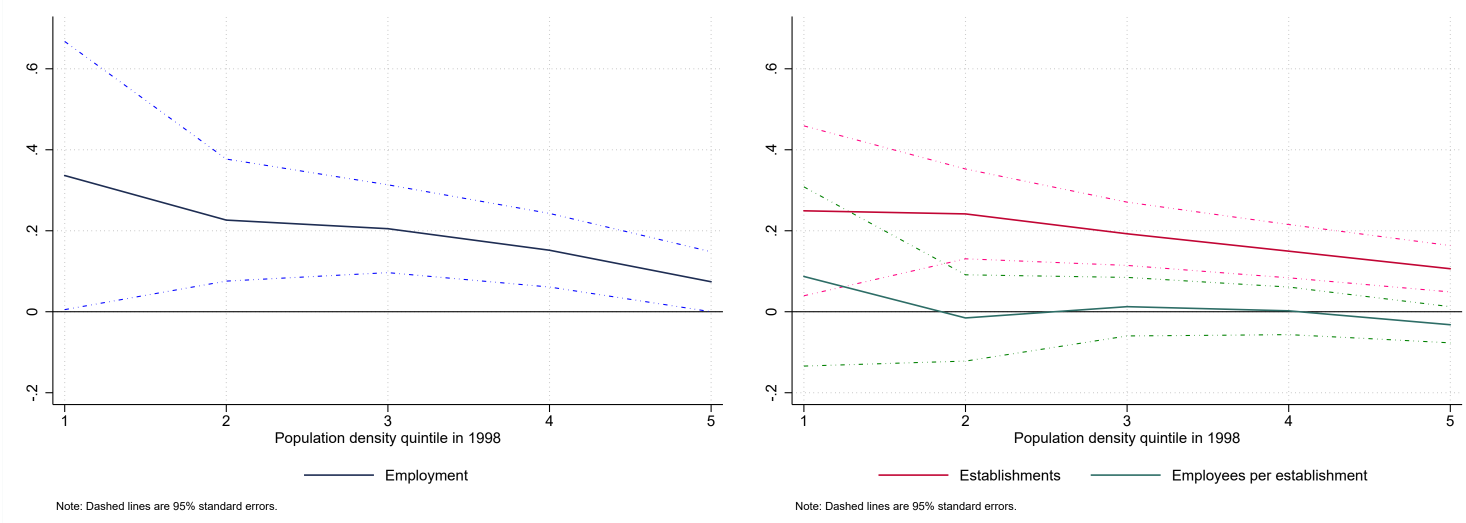

Figure 5: Slope of interaction of population and frequency (IV)

Table 8: **Number of employees per establishment and frequency of purchase (2SLS)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.286***	0.314***	0.299***	0.270***	0.421***
	(0.044)	(0.038)	(0.036)	(0.055)	(0.047)
Log population × log frequency	-0.031	-0.024	-0.028	-0.026	-0.032
	(0.026)	(0.026)	(0.023)	(0.023)	(0.023)
Log population × CEX expenditure share	-0.244	-0.800**	-0.516	-0.489	-0.504
	(0.412)	(0.400)	(0.353)	(0.353)	(0.352)
Log income per capita	0.174***	0.228***	0.203***	0.330***	0.076
	(0.059)	(0.064)	(0.058)	(0.127)	(0.117)
Log land area	0.004	0.001	0.002	0.023	-0.058**
	(0.006)	(0.006)	(0.006)	(0.015)	(0.026)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.26	0.27	0.26	0.23	0.19
N	60,413	60,923	121,336	121,336	121,336
S.W. F stat: Log Population	28.03	43.35	37.78	11.92	20.14
S.W. F stat: Log Population × log frequency	20.44	24.9	25.74	13.43	16.28
S.W. F stat: Log Population × CEX expenditure share	12.81	15.89	15.29	16.55	14.88

Standard errors clustered at county level in parentheses.

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

local economic outcomes. In sectors with high storage costs, consumers should be more willing to trade off larger batches with frequent trips, but to do so they would choose to travel shorter distances. An implication is that, in response to a larger population, firms in high storage-cost sectors face a relatively stronger incentive to increase production by economizing on land; hence, local employment should increase relatively more. We proxy for storability/durability using the frequency of transactions in our credit card data: our simple model suggests that storage costs induce variation in such frequency, conditional on measures of overall consumer expenditure. We find support for the relevance of consumer mobility in shaping the local structure of consumption industries. This support appears present both in a subset of counties where endogeneity concerns may be more muted, and across all counties when an instrumental variable strategy offers a way of addressing such endogeneity. Employment responds via a denser network of local suppliers, which reduces the average distance between consumers and establishments within a county. This channel of adjustment is compatible with the importance of time savings for consumers in sectors more frequently transacted. The economic significance of storability/durability appears roughly equal to the significance of budget share variations in influencing local equilibrium outcomes.

5 Extensions, Robustness and Limitations

In this section, we summarize ancillary results and robustness exercises reported in the Appendix; we also emphasize some limitations of our analysis.

Spillovers into neighboring counties

We have argued that, in response to a larger local population, local employment is relatively larger in high-frequency sectors because the demand for those sectors is more spatially concentrated. One might also think of examining the consequences of a larger *local* population on employment and establishments in *neighboring* counties. If local demand in higher frequency sectors is relatively more concentrated, it should spill relatively less into neighboring counties; however, a larger CEX expenditure share should still imply larger neighboring employment, everything else equal. Testing these implications is harder for a number of reasons: for example, the shape of the transportation infrastructure may tilt demand flows towards a particular subset of neighboring counties, or even beyond immediate neighbors. Nonetheless, we analyze the response of neighbors' outcomes in Appendix E.2. In particular, we construct employment, establishments, and employees per establishments in counties whose centroid is within 120 km from the focal county. Tables E.4-E.6 then replicate Tables 6-8 using these neighboring counties outcomes as dependent variables, and controlling for neighboring county observables. We find that a larger CEX expenditure share on a sector still increases employment in neighboring counties relatively more. On the other hand, in response to a larger *local* population, employment grows relatively less in high-frequency sectors (where *local* demand should be more spatially concentrated) than in low-frequency sectors (where local demand should be more likely to spill over). Consistent with a consumer mobility-based explanation, high-frequency sectors still see relatively more establishments.

Density and Sorting on Preferences

Our instrumental variable is valid if aquifers affect employment only through population. For fixed land area, however, the presence of aquifers increases both population and density by construction. Hence, in principle our exercise may pick up a comparison between a large (and high-density) place to a small (and low-density) place. It could be the case that density has an independent differential effect across sectors on top of population. For example, one might imagine that aquifers are associated with older cities which are now denser, scarcer in space and where both residential and commercial rents are high; it might be the case that the type of people that sort into these places have a stronger preference for high-frequency sectors with respect to people that demand more space and possibly sort into newer and maybe less dense cities. If this is true, places with higher density may attract high-frequency types of economic activity relatively more.

A series of considerations indicate that these concerns may be of limited practical importance. First, the battery of regressions in Table 6-8 include a set of increasingly narrow geographic fixed effects: in the strictest specification, we are comparing counties within the same commuting zone, rather than big and dense commuting zones vs. small and sparse ones. Second, if the story we described was empirically relevant, any heterogeneity in the impact of density would lead the coefficient on the interaction term to *grow* as we progressively include denser counties; however, Figure 5 shows that, if anything, the coefficient becomes smaller. Third, we can run regressions that directly control for a differential effect of income-varying preferences or of density. In Appendix E.3, we report more details about these exercises. As a summary, we first control for an interaction between income per capita and log frequency of transactions to allow for heterogeneity in the role of income on local demand. Table E.7 shows that higher income is associated to a relatively smaller employment in high-frequency sectors, while the coefficient on population times frequency is little affected. In Table E.8 we then control for density, and its interaction with frequency, directly. When we change population controlling for density, we are comparing, say, a small sprawling place to a big sprawling place. The interaction between density and frequency is *negative*, that is, a denser place has relatively less employment in high-frequency sectors than in low-frequency ones; this negative sign is consistent with the evidence presented in Figure 5, that shows a declining slope on the interaction as denser counties are included in the sample. Importantly, the coefficient on the interaction between population and frequency stays positive and significant.

Fixed costs. In Tables 6-8, we have argued that an exogenous increase in population tends to generate demand that is more geographically concentrated for high storage-cost sectors than low storage-cost ones. Heterogeneous fixed costs across industries, however, may also affect the density of establishments and hence confound our estimates.

Two considerations support the view that these issues are not a primary concern. Similar to storage costs, direct observations of fixed costs are hard to obtain. However, a reasonable proxy is the economy-wide ratio of total employment to total establishments in a sector-year, i.e., the average establishment size. If fixed costs are high, increasing returns to scale are more important, and we should expect a higher employees-to-establishment ratio. With this measure in hand, we first find that the correlation between the log average frequency of transactions and the log proxy for fixed costs across sectors is only 0.12, and statistically indistinguishable from zero: hence, it is not empirically true that low fixed cost sectors are those where transactions are more frequent. Second, we can explore the sensitivity of our analysis to the interaction of fixed costs with population. This new interaction variable is again instrumented with the interaction between fixed costs and county geological composition. Table E.9 in Appendix E.4 replicates the most conservative specifications in columns (5) and (6) for Tables 6-8. The coefficient on the interaction between frequency and population stays positive and of very similar magnitude. The coefficients on the interaction with the proxy for fixed costs are small and insignificant. Overall, we read these results as further evidence consistent with a role of consumers' mobility on local economic outcomes.

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

Trip chaining. It is natural to think that one way in which consumers optimize their shopping behavior is to make a number of possibly unrelated purchases on a single trip to a commercial area. For example, Shoag and Veuger (2017) document positive externalities of "big-box" stores on neighboring businesses via the increased local foot traffic. Our data is unfortunately too coarse to speak to that aspect: out of all account-transaction dates in our data, only 25% have more than one transaction per day, and less than 1% have at least 5 transactions; of the cases in which there is more than one transaction, 80% are multiple transactions occurring in the same broad sector. How would our results be impacted if this was the predominant behavior of consumers? Suppose that consumers always travel to one mall and buy food every trip, but apparel every four trips. In that case, we would expect to see no relation between gravity and the frequency of transactions, since the frequency of purchase differs, while the distance stays constant. More importantly, the frequency of transactions in the credit card data would less likely predict heterogeneity in the impact of population on store density. The fact that we see at least some impact is indicative that trip chaining is not the only relevant feature of the data.[48] As above, however, we emphasize that our conclusions should be considered accounting for this limitation.

6 Conclusion

In developed countries, a large fraction of economic activity in terms of employment and GDP is involved in the delivery of final consumption. The local composition of these industries matters, among other things, for the relative attractiveness of different places and the type of skills that they demand. These outcomes, in turn, help to shape the economic success or decline of different areas. In this paper, we study the determinants of the local structure of consumption industries, i.e., the local employment, the number of establishments, and establishment size for industries delivering final consumption. The significant travel and time investment that consumers incur for their purchases requires attention to the main features of consumer mobility. In our work, we characterize these features across sectors for the first time. We then argue that consumer mobility appears to influence the local structure of consumption industries in an economically intuitive way: where travel frictions are expected to matter relatively more, employment grows relatively more concentrated, and establishments are relatively denser, in response to a larger local population. The economic significance of storability/durability appears approximately as large as that associated with variation in consumers' expenditure shares.

One could see our findings as a local empirical analog of the classic proximity-concentration trade-off (PCT) in the foreign direct investment (FDI) literature.[49] Firms in a location choose whether to

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

[49] see for example Horstmann and Markusen (1992); or Brainard (1997).

serve customers in a neighboring location by expecting them to travel ("export" in PCT) or by opening a new establishment closer to them ("FDI" in PCT); in sectors where the frequency of transactions is high (high "transportation costs" in PCT), demand is more localized, and opening a new establishment, i.e. the "FDI" option, becomes more attractive. In this sense, low storability/durability, high-frequency industries are less "tradeable." This analogy is useful when considering the expansion strategy of firms supplying services: whether within or across countries, some of these firms need to decide the number and location of establishments to open in order to serve a particular geographical market. We suggest that the geographical scope of an establishment market is in part related to the interplay between durability/storability and consumers' mobility.

Interestingly, even in the set of consumption industries that we analyze, we find a classic fact that associates more trade and openness to higher average wages:[50] the correlation between annual log wages in the 2007 County Business Patterns and log frequency of transactions is -0.65 across our 21 sectors.

Our results are subject to caveats that arise from the limited nature of our data and from the attempt to bring under a unified logic consumption markets that are potentially quite different. On the other hand, our findings describe broad patterns of consumer behavior for a large portion of economic activity in modern economies. Taken together, they suggest the importance of consumer mobility, as well as of factors that facilitate or hinder it, in shaping important local outcomes. We hope this work will stimulate further research in this direction.

[50] See, for example, Bernard and Jensen (1999), Jensen (2011), or Gervais and Jensen (2019).

References

[1] Abdel-Rahman, H. M. and A. Anas (2004). "Evidence on the Nature and Sources of Agglomeration Economies," Handbook of Urban and Regional Economics, Chapter 52, Elsevier.

[2] Alchian, A., and W. Allen (1964). "University Economics". Belmont, Calif.: Wadsworth.

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

[4] 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.

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

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

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

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

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

[10] Bernard, A. and J.B. Jensen, 1999. "Exceptional Exporter Performance: cause, effect, or both?" Journal of International Economics, Vol. 47:1, pp. 1-25.

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

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

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

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

[15] Chevalier, J. and A. Kashyap (forthcoming). "Best Prices: Price Discrimination and Consumer Substitution," American Economic Journal: Economic Policy, Vol 11:1, pp. 126-59.

[16] Christaller, Walter (1933). “Die zentralen Orte in Süddeutschland.” Gustav Fischer, Jena.

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

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

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

[20] Couture, V. (2016). “Valuing the Consumption Benefits of Urban Density,” Working Paper.

[21] Couture, V., C. Gaubert, J. Handbury and E. Hurst (2019). “Income Growth and the Distributional Effects of Urban Spatial Sorting,” Working Paper, July.

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

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

[24] Davis, D. and J. Dingel (2019). “The Comparative Advantage of Cities,” mimeograph, The University of Chicago.

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

[26] Diamond, R. (2016). “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” American Economic Review, 106(3), pp. 479-524.

[27] Dolfen, P., L. Einav, P. Klenow, B. Klopack, J. Levin, L. Levin, and W. Best (2019). “Assessing The Gains from E-Commerce,” Working Paper.

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

[29] Duranton, G. and D. Puga (2001). “Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products,” American Economic Review, Vol. 91:5, pp. 1454-1477.

[30] Duranton, G. and D. Puga (2005).“From sectoral to functional urban specialisation,” Journal of Urban Economics, Vol. 57:2, pp. 343-370.

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

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

[33] Eckert, F.P. (2019). "Growing Apart: Tradeable Services and the Fragmentation of the U.S. Economy," mimeograph, Yale University.

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

[35] Ellison, G., Glaeser and W. Kerr (2010). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," American Economic Review, Vol. 100:3, p. 1195-1213.

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

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

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

[39] Gervais, A. and J.B. Jensen (2019). "The Tradability of Services: Geographic Concentration and Trade Costs," Journal of International Economics, Vol. 118:C, pp. 331-350.

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

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

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

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

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

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

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

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

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

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

[50] Hsu, , W. (2012). "Central Place Theory and City Size Distribution," The Economic Journal, Vol 122 (September), pp. 903-932.

[51] Hsu, W., T. Holmes and F. Morgan (2014). "Optimal City Hierarchy: A dynamic Programming Approach to Central Place Theory„" Journal of Economic Theory, Vol. 154, pp. 245-273.

[52] Jensen, J. B. and L. Kletzer (2010). "Measuring Tradable Services and the Task Content of Offshorable Services Jobs," in "Labor in the New Economy," NBER.

[53] Jensen, J.B. (2011). "Global Trade in Services: Fear, Facts, and Offshoring," Peterson Institute for International Economics Press.

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

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

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

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

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

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

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

[61] 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.

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

[63] Redding, S. and E. Rossi-Hansberg (2018). “Quantitative Spatial Economics,” Annual Review of Economics, Vol. 9:21-58, pp.21-58.

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

[65] Rosenthal, S. and W. Strange (2004). “Evidence on the Nature and Sources of Agglomeration Economies,” Handbook of Urban and Regional Economics, Chapter 49. Elsevier.

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

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

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

[69] 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.

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

Web Appendix for "Consumer Mobility and The Local Structure of Consumption Industries" (Not for Publication)

Sumit Agarwal
National University of Singapore

J. Bradford Jensen
Georgetown University

Ferdinando Monte
Georgetown University

A Data Processing

A.1 Merchant codes

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

A.2 Transaction data

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

A.3 Account data

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

A.4 Matching transactions and account data

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

A.5 Extracting merchant location name

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

We first extract the merchant state. The state of the merchant is typically located at the end of the full merchant name. We extract the last two characters of the merchant name string if the last three start with a space. Only 1,588 transactions do not meet this requirement: in most cases, the last two letters still represent a state (or a foreign country), but we won't be able to rule out false positives. We match these states with a list of U.S. states and country abbreviations to verify that we have extracted U.S. states. We match only 52% of the 1,588 thousand problematic observations, and more than 98% of the other transactions. Keeping only transactions where a U.S. state could be identified leaves us with 2,106,552 observations.

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

We then mark transactions of common online providers[51] and find the words "Online", "On Line", ".com", ".net" in 100,265 observations. We mark observations where the final part of the string before the state is a phone number – these are typically online stores – and find 188,316 of them. We are left with 1,901,658 transactions that may contain city names, 90% of those for which a state name could be found, for 73,385 unique accounts. Note that the largest contributor to the drop in observations is transactions with a phone number rather than a location at the end of the merchant name. We will attempt to match this list of location names with a list of U.S. city and place names from the U.S. Census. Before turning to the different steps in that match, we will discuss briefly how we recover the list of cities.

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

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

While FIPS codes are unique, our match to merchants will be on a location name. Hence it may happen that within the list, we have more than one record with the same name (for example, we may have "Mountain View city" and "Mountain View, CDP"). In those cases, we attribute to a name the coordinates with the highest population in 2000.[52]

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

We attempt to find the name of a city in four passes. First, we match the location name and state identified above with the list of U.S. Places. We immediately find a match for 1,454,166 out of the 1,901,658 we intend to match, 76% of our observations. Out of the 447,492 transaction with no match, 122,737 have names and states that match the MCD list. We assign "match quality" equal to 10 to those transactions matched at this first pass. We have 324,755 transactions with no location information (about 17% of the transactions) that we cannot match exactly.

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

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

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

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

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

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

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

B Theoretical derivations

B.1 Equilibrium Price

The generic solution to (8) is

$$p(j) = c_1 \exp\{\alpha j\} + c_2 \exp\{-\alpha j\} \tag{B.1}$$

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

Using (B.1), it follows that

$$-c_1 \exp\{\alpha j_{\max}\} = c_2 \exp\{-\alpha j_{\max}\} \tag{B.2}$$

Using the same information in the implicit function for the distance traveled (eq. (6)), when $t = 1 \Longrightarrow j(1) = j_{\max}$, and hence,

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

We can substitute (B.2) in the last equation to obtain,

$$\begin{aligned} \left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} &= c_2 \exp\{-\alpha j_{\max}\} + c_2 \exp\{-\alpha j_{\max}\} \\ c_2 &= \frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{\alpha j_{\max}\} \end{aligned} \tag{B.4}$$

and hence

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

We can then rewrite the price function as

$$\begin{aligned} p(j) &= -\frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{-\alpha j_{\max}\} \exp\{\alpha j\} + \frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \exp\{\alpha j_{\max}\} \exp\{-\alpha j\} = \\ &= \frac{1}{2}\left(\frac{2g}{\bar{q}}\right)^{1/2} \frac{1}{\alpha} \left[\exp\{\alpha (j_{\max} - j)\} - \exp\{-\alpha (j_{\max} - j)\}\right] \end{aligned} \tag{B.5}$$

and the implicit equation for distance,

$$t = \frac{1}{2}\left[\exp\{\alpha (j_{\max} - j)\} + \exp\{-\alpha (j_{\max} - j)\}\right] \tag{B.6}$$

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

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

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

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

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

B.2 Aggregate Implications

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

Proof. Recall from (9) that

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

Differentiating with respect to j, we have

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

The expenditure function $X(j)$ has slope

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

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

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

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

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

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

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

which is again more negative if g is higher. ■

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

Proof. Since the frequency of trips is $\bar{q}/z$, we consider the behavior of the batch size. We show that agents buy smaller batches as g grows, implying they travel more frequently. From (4) and using the functional form assumptions, the batch size is

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

where we have evaluated the batch at the optimal distance for agent t, $j(t)$, and we have made the dependence of the travel function on the parameter g explicit. As g grows, the optimal batch for given distance $j(t;g)$ shrinks via the denominator. Also, the optimal distance traveled $j(t;g)$ falls with g for every agent. To see this, recall that $j(t)$ is implicitly defined by (B.6). Totally differentiating with respect to j and t,

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

where we have used $p''(j) = \alpha^2 p(j)$ from (8) and the definition of α. Differentiating with respect to t, one can verify that $j(t)$ is always convex:

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

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

Lemma 4 . *The equilibrium total employment is*

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

■

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

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

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

Lemma 5 . *The average distance at which output is produced is*

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

Proof. Using the expression for output (2),

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

■

C Additional Empirical Results: Descriptive Patterns

C.1 Summary Statistics by state

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

C.2 Frequent users

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

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

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

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

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

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

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

Table C.3: Locations available and locations visited

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

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

C.3 Percentiles of distances traveled

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

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

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

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

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

C.4 Gravity over all distances

In Figure C.1, we estimate Equation (16) including origin-destination pairs at progressively longer distances. Specifically, we split all the (h, s) pairs in 20 quantiles of distances, and estimate it using only the first group, then only the first two, and so on, up to the whole set of observations. Figure C.1 shows the coefficient on log distance. As one can see, changes of around +/- 30% in the 120 km cutoff (from 80 km to 160 km) only imply a variation in the gravity coefficient of around 0.1: hence, around our cutoff distance, the overall gravity slope is not particularly sensitive to the specific cutoff value. Different sectors are more or less represented at different distances (see also Tables C.4 and C.5), implying that the coefficient δ varies.

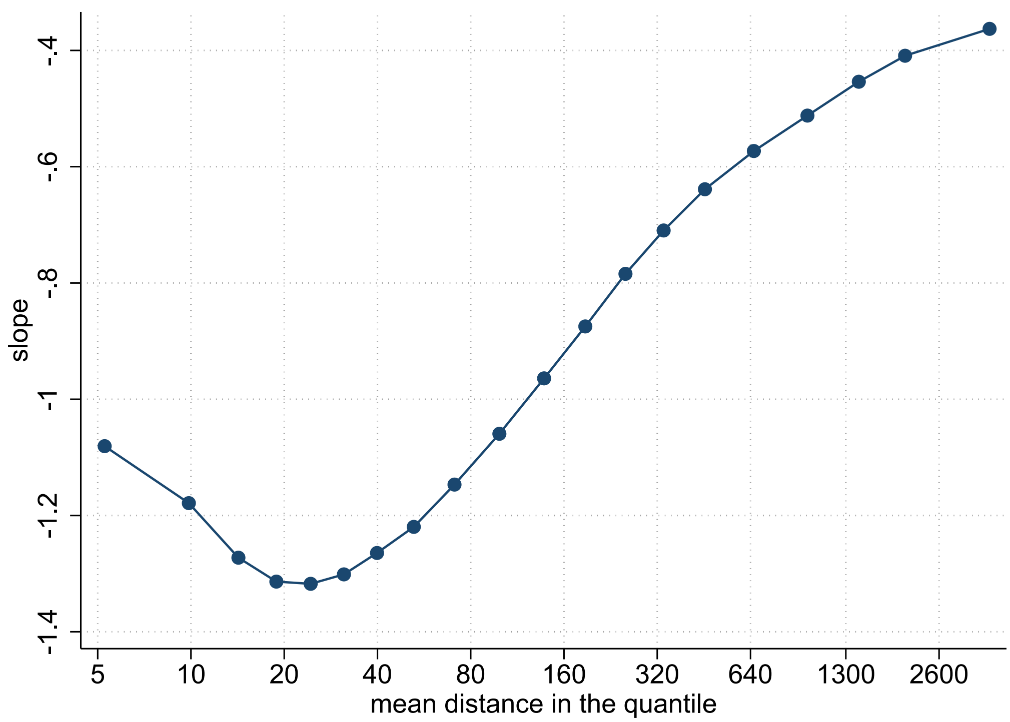

Figure C.1: Gravity in Expenditure

C.5 Margins decomposition

Tables C.6 shows the actual values of the account and expenditure margin with associated p-values for the margins decomposition associated with Equation (15); Table C.7 shows the actual values of the account and expenditure margin with associated p-values associated with Equation (16).

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

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

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

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

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

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

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

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

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

C.6 Gravity and the frequency of transactions

These figures show further robustness on the relation between gravity and the frequency of transactions. Figure C.2 shows the correspondent of Figure 3 using all coefficients, not just the ones significantly different from zero; one can clearly note the outlier "Durable Goods" in the top-left part of the graph. Figure C.3 uses the strength of gravity as measured by regression (16) using all estimated slopes. We have also experimented with an alternative measure of frequency that gives more weight to users that spend more overall, with essentially identical results.

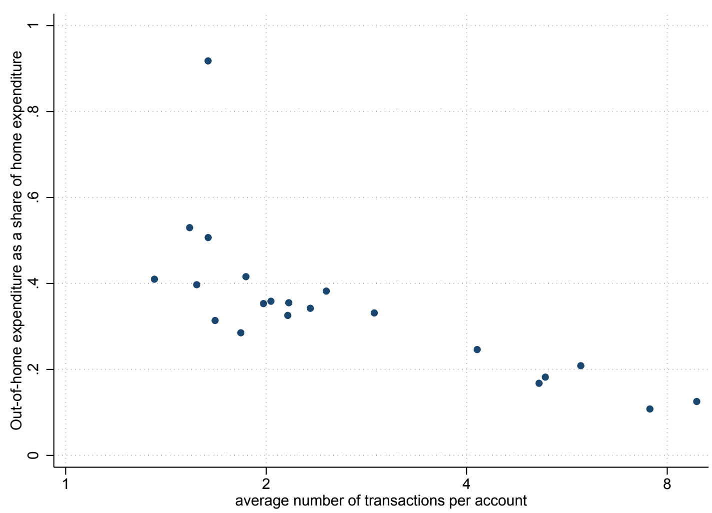

Figure C.2: Drop in expenditure out of home (all coefficients)

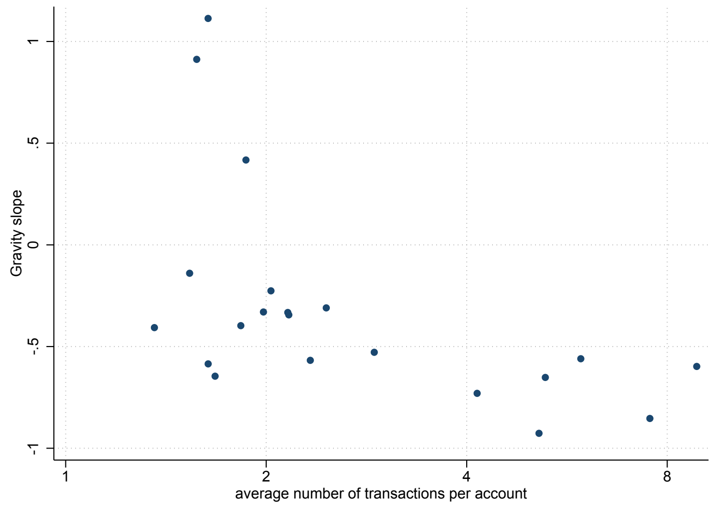

Figure C.3: Gravity and frequency of transactions (all slopes)

D Additional Empirical Results: Individual Level Responses

The analysis in Section 3 has shown that the strength of gravity varies by sector with the frequency of travel, which we interpret as related to sector-level characteristics like storability and durability. In this Appendix, we ask if, for a fixed network of suppliers, individual characteristics predict additional sector-level heterogeneity in travel behavior. We first explore whether individual heterogeneity affects geographic patterns of purchase across sectors. We then compare the behavior of the same individual under two travel cost regimes, and ask whether the travel response to this shock across sectors is a function of individual characteristics.

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

D.1 The Role of Individual Characteristics

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

$$E\left[out_{i,s}|n_{i,s}, x_i, \tilde{x}_i^s\right] = \exp\left\{\alpha + \beta_0 \cdot n_{i,s} + \gamma_0' \cdot x_i + \gamma_1' \tilde{x}_i^s + \xi_i + \zeta_s\right\} \tag{D.1}$$

In this expression, $n_{i,s}$ is the total number of transactions for account i in sector s; x_i are individual log age and log income; $\tilde{x}_i^s$ are a set of interactions between those individual variables and three sector-level characteristics: the frequency of transactions, the CEX expenditure share, and the U.S.-wide average number of employees per establishment from the County Business Patterns in 1998-2007; and ξ_i and ζ_s are vectors of individual and sector dummies. Note that since we are controlling for the total number of transactions, this analysis effectively examines the response of the geographical composition of purchases to different covariates; also note that shape of the geographical units of observations plays a limited role since our dependent variable sums across all locations other than the residence location.

Our coefficients of interest are the interactions of demographic characteristics with the average frequency of transactions. Table D.1 shows the results of this estimation. In all specifications, we cluster standard errors at the individual level to allow for correlation in the number of transactions out-of-home within individuals across sectors. All the models include sector-fixed effects. Column (1) shows the baseline elements of our regression, that compares individuals living in the same zip code. The expected number of transactions out-of-home in the sample period increases by 2.2 percentage points for an individual with one additional transaction.[53] In column (2), we interact log age and log income with two sector characteristics, the average frequency of transactions in a sector, and the CEX expenditure share. We find that these two sector characteristics do not interact significantly with age. We find however that income is a determinant of the geographical composition of trips, and that the role of income varies as a function of the proxy for storability/durability. Moving from a sector at the 10th percentile of frequency (Miscellaneous Services), to a sector at the 90th percentile of frequency (Eating and Drinking Places), the elasticity of the number of out-of-home transactions to a 10% increase in income falls by 2.35 percentage points. This variation occurs controlling for the CEX expenditure share in the sector. In comparison, a 10th-90th percentile move (from Miscellaneous Services to Eating and Drinking Places) in the CEX share implies a drop of 1.55 percentage points in the same elasticity.

The coefficient on these interactions stays stable in columns (3), where we control for individual fixed-effects absorbing time-invariant person-level characteristics (e.g., wealth, education, precise residence location, overall use of credit cards); and in column 4, where we additionally interact individual level variables with the national average size of establishments, a proxy for sector level fixed-costs.

In unreported results, we have explicitly modeled overdispersion and estimated a Negative Binomial Model: when we do that, we find that the patterns of significance are unchanged. We have also experimented with alternative clusterings of the standard errors: for example, we have split individuals in 5 income quintiles and 5 age quintiles; we then have formed indicators for the 525 combinations of age group times income group time sector and clustered our standard errors using this alternative categorization. The estimates of the interaction coefficients maintain the same patterns of significance.

D.2 The Effect of Rain

In the section above, we have asked whether individual characteristics shape the geographic distribution of transactions across sectors. Such an exercise leveraged cross-sectoral variation across individuals, controlling as possible for individual characteristics. In this section, we consider the same individual under

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

two travel cost regimes: we examine whether the response of the geographic distribution of transactions across sectors to a travel cost shock is different for, say, high income vs. low income individuals.

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

We then construct an extended dataset, starting from the analysis in the subsection above, where for each individual we count $out_{i,s,r}$, the number of transactions out-of-home by sector during rainy ($r = 1$) and non-rainy days ($r = 0$). Our interest is in a "triple-interaction" variable: we compare the response of, say, low income individuals and high income individuals to the same travel cost shock, and ask whether there are differences in the sectoral heterogeneity of their responses.

In particular, we estimate Poisson models where the mean number of out–of–home transactions for individual i in sector s under rain conditions r, $out_{i,s,r}$, takes the general form:

$$\begin{aligned} E\left[out_{i,s,r}|r, x_i, \bar{f}_s, \delta^i, \delta^s, \delta^r\right] &= \exp\left\{\alpha_0 + \alpha_{0,r} \cdot I_{r=1} + \beta_0 \cdot n_{i,s,r} + \beta_{0,r} \cdot I_{r=1} \cdot n_{i,s,r} + \xi_i + \zeta_s + I_{r=1} \cdot \sigma_s + \right. \\ &\quad \left. \gamma_0' \cdot x_i + \gamma_{0,r}' \cdot I_{r=1} \cdot x_i + \gamma_1' \tilde{x}_i^s + \gamma_{1,r}' \cdot I_{r=1} \cdot \tilde{x}_i^s \right\} \end{aligned}$$

In this expression, $n_{i,s,r}$ is the total number of transactions recorded for account i by sector s and weather status r; $I_{r=1}$ is a dummy equal to 1 for observations pertaining to rainy days and zero otherwise; ξ_i and ζ_s are individual and sector dummies; σ_s are the same three sector characteristics as above, to be interacted with the rain dummy; x_i are log age and log income of the individual; $\tilde{x}_{i,}^s$ are a set of double interactions of individual and sector-level characteristics.

The results of this analysis are reported in Table D.2, which mimics the structure of Table D.1 above: all columns include sector fixed effects; Column (1) includes only the level variables and zip code fixed effects; column (2) includes all the rain-sector-individual double interactions necessary to examine the role of the frequency of transactions and CEX expenditure; column (3) includes all the relevant triple interactions; column 4 replaces zip code fixed-effects with individual fixed-effects; and column (5) includes the triple interactions with our proxy for fixed costs, plus the remaining necessary double interactions. Again, all standard errors are clustered at individual level. Our coefficient of interests are on rain × log frequency and on the triple interaction between rain, log income, and log frequency of transactions. These coefficients are fairly stable across specifications. The positive coefficient on rain × log frequency indicates that for a relatively low income individual, transactions in low-frequency sectors become relatively more local than those in high-frequency sectors; this relation, however, becomes flatter for higher income people.

In unreported results, we replicate our robustness exercises described above for a Negative Binomial Model, and the same alternative clustering of standard errors: significance patterns are preserved.

These results suggest that the response of the geographic distribution of transactions to a travel cost shock is different, across sectors, as a function of individual characteristics. Since the network of suppliers is stable, and our "cost regimes" comparison keeps individual unobservables fixed, we interpret these results as further evidence that consumers actively choose the distance traveled in a way that is related to our proxy for storability/durability.

Table D.1: **The role of individual heterogeneity (poisson regression)**

	(1)	(2)	(3)	(4)
Dependent Variable:	Number of transactions out of residence			
Log income	-0.067	0.268***		
	(0.044)	(0.058)		
Log age	-0.048	0.078		
	(0.084)	(0.129)		
Number of transactions	0.022***	0.022***	0.023***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
Log income × log frequency of transactions		-0.177***	-0.173***	-0.174***
		(0.024)	(0.024)	(0.023)
Log age × log frequency of transactions		-0.031	-0.029	-0.023
		(0.055)	(0.055)	(0.055)
Log income × CEX expenditure share		-0.629**	-0.614**	-0.639**
		(0.295)	(0.296)	(0.299)
Log age × CEX expenditure share		-0.952	-0.946	-0.959
		(0.691)	(0.699)	(0.692)
Log income × log of employees per store				-0.051***
				(0.013)
Log age × log of employees per store				0.117***
				(0.030)
Observations	28,959	28,959	28,959	28,959
Sector fixed effects	Yes	Yes	Yes	Yes
ZIP code fixed effects	Yes	Yes	No	No
Individual fixed effects	No	No	Yes	Yes
Pseudo R-Square	.74	.74	.74	.74

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

Table D.2: **The effect of rain (poisson regression)**

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Number of transactions out of residence				
Log age	-0.016	0.098	0.155		
	(0.078)	(0.126)	(0.138)		
Log income	-0.059	0.289***	0.268***		
	(0.041)	(0.058)	(0.062)		
Number of transactions	0.034***	0.035***	0.035***	0.036***	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Rain dummy	-0.221***	-0.876***	-0.916	-0.966	-1.130
	(0.023)	(0.270)	(0.673)	(0.670)	(0.916)
Rain dummy × log frequency of transactions		0.358***	0.699**	0.705**	0.765**
		(0.034)	(0.348)	(0.349)	(0.346)
Rain dummy × log income		0.001	0.058	0.059	0.088
		(0.019)	(0.052)	(0.051)	(0.065)
Rain dummy × log age		-0.021	-0.183	-0.172	-0.233
		(0.037)	(0.127)	(0.127)	(0.169)
Log income × log frequency of transactions		-0.188***	-0.168***	-0.165***	-0.164***
		(0.024)	(0.027)	(0.027)	(0.027)
Log age × log frequency of transactions		-0.052	-0.077	-0.076	-0.064
		(0.055)	(0.065)	(0.065)	(0.065)
Rain dummy × CEX expenditure share		1.923***	-5.443	-4.561	-5.088
		(0.229)	(5.583)	(5.542)	(5.519)
Log age × CEX expenditure share		-0.339	-0.568	-0.472	-0.525
		(0.705)	(0.910)	(0.913)	(0.901)
Log income × CEX expenditure share		-0.592**	-0.767**	-0.726**	-0.762**
		(0.282)	(0.362)	(0.360)	(0.364)
Rain × log income × log frequency of transactions			-0.054**	-0.053**	-0.058**
			(0.026)	(0.026)	(0.026)
Rain × log age × log frequency of transactions			0.070	0.069	0.068
			(0.072)	(0.072)	(0.072)
Rain × log income × CEX expenditure share			0.456	0.424	0.466
			(0.388)	(0.386)	(0.385)
Rain × log age × CEX expenditure share			0.604	0.458	0.472
			(0.924)	(0.926)	(0.912)
Rain dummy × log of employees per store					0.044
					(0.206)
Log age × log of employees per store					0.139***
					(0.037)
Log income × log of employees per store					-0.045***
					(0.015)
Rain × log income × log of employees per store					-0.009
					(0.014)
Rain × log age × log of employees per store					0.021
					(0.040)
Observations	57,918	57,918	57,918	57,918	57,918
Sector fixed effects	Yes	Yes	Yes	Yes	Yes
ZIP code fixed effects	Yes	Yes	Yes	No	No
Individual fixed effects	No	No	No	Yes	Yes
Pseudo R-Square	.68	.68	.68	.69	.69

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

E Additional Empirical Results: County-Level Analysis

E.1 Weak Instruments

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

Table E.1-E.3 report the corresponding LIML estimates of Tables 6-8. Coefficients are broadly in line, and standard errors essentially unchanged. These findings further help alleviate potential concerns about weak instruments.

Table E.1: **Local employment and frequency of purchase (LIML)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log employment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.733***	0.884***	0.808***	0.795***	1.208***
	(0.105)	(0.075)	(0.085)	(0.129)	(0.086)
Log population × log frequency	0.130**	0.074*	0.101**	0.090**	0.076**
	(0.052)	(0.041)	(0.043)	(0.042)	(0.038)
Log population × CEX expenditure share	0.828	0.993*	0.922**	0.889**	0.792*
	(0.530)	(0.513)	(0.458)	(0.451)	(0.438)
Log income per capita	1.643***	1.606***	1.633***	1.785***	1.018***
	(0.149)	(0.137)	(0.137)	(0.303)	(0.205)
Log land area	0.103***	0.136***	0.120***	0.185***	-0.023
	(0.013)	(0.011)	(0.011)	(0.035)	(0.044)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.76	0.79	0.78	0.72	0.69
N	60,413	60,923	121,336	121,336	121,336
S.W. F stat: Log Population	28.03	43.35	37.78	11.92	20.14
S.W. F stat: Log Population × log frequency	20.44	24.9	25.74	13.43	16.28
S.W. F stat: Log Population × CEX expenditure share	12.81	15.89	15.29	16.55	14.88

Standard errors clustered at county level in parentheses.

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

Table E.2: **Number of establishments and frequency of purchase (LIML)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.417***	0.556***	0.490***	0.500***	0.771***
	(0.099)	(0.065)	(0.078)	(0.108)	(0.066)
Log population × log frequency	0.174***	0.105***	0.137***	0.122***	0.117***
	(0.046)	(0.033)	(0.037)	(0.035)	(0.033)
Log population × CEX expenditure share	1.123***	1.864***	1.501***	1.424***	1.348***
	(0.337)	(0.371)	(0.308)	(0.291)	(0.286)
Log income per capita	1.501***	1.389***	1.449***	1.502***	0.955***
	(0.136)	(0.105)	(0.114)	(0.246)	(0.145)
Log land area	0.099***	0.135***	0.118***	0.166***	0.038
	(0.012)	(0.008)	(0.009)	(0.029)	(0.031)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.80	0.83	0.82	0.78	0.76
N	60,413	60,923	121,336	121,336	121,336
S.W. F stat: Log Population	28.03	43.35	37.78	11.92	20.14
S.W. F stat: Log Population × log frequency	20.44	24.9	25.74	13.43	16.28
S.W. F stat: Log Population × CEX expenditure share	12.81	15.89	15.29	16.55	14.88

Standard errors clustered at county level in parentheses.

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

Table E.3: **Number of employees per establishment and frequency of purchase (LIML)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	county-sector log number of employees per establishment				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.286***	0.314***	0.299***	0.269***	0.422***
	(0.044)	(0.038)	(0.037)	(0.056)	(0.048)
Log population × log frequency	-0.031	-0.024	-0.028	-0.026	-0.032
	(0.026)	(0.026)	(0.023)	(0.024)	(0.023)
Log population × CEX expenditure share	-0.245	-0.803**	-0.518	-0.492	-0.506
	(0.415)	(0.403)	(0.355)	(0.358)	(0.354)
Log income per capita	0.174***	0.228***	0.203***	0.335**	0.075
	(0.059)	(0.064)	(0.058)	(0.130)	(0.118)
Log land area	0.004	0.001	0.002	0.024	-0.059**
	(0.006)	(0.006)	(0.006)	(0.015)	(0.026)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.26	0.27	0.26	0.23	0.19
N	60,413	60,923	121,336	121,336	121,336
S.W. F stat: Log Population	28.03	43.35	37.78	11.92	20.14
S.W. F stat: Log Population × log frequency	20.44	24.9	25.74	13.43	16.28
S.W. F stat: Log Population × CEX expenditure share	12.81	15.89	15.29	16.55	14.88

Standard errors clustered at county level in parentheses.

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

E.2 Spillovers into Neighboring Counties

In this section, we report the results of running 2SLS specifications of eq. (19) on neigboring counties' outcome. In particular, for each focal county, we identify the set of counties whose centroid is less than 120 km away from the focal county's centroid. For this set of counties, we compute total employment, total number of establishments, employees per establishment, total population, total land area and average income per capita. Tables E.4-E.6 below then replicate Tables 6-8 with these alternative outcomes.

Table E.4: **Local employment around and frequency of purchase (2SLS)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	neighboring counties-sector log employment				
Sample years :	07	98	98,07	98,07	98,07
Log population	-0.247**	-0.063	-0.169*	-0.169*	0.038*
	(0.117)	(0.076)	(0.095)	(0.095)	(0.022)
Log population × log frequency	-0.080***	-0.078***	-0.079***	-0.079***	-0.087***
	(0.017)	(0.015)	(0.016)	(0.016)	(0.016)
Log population × CEX expenditure share	1.295***	0.738***	1.022***	1.022***	0.954***
	(0.232)	(0.156)	(0.183)	(0.183)	(0.173)
Log land area around	-0.119	0.026	-0.058	-0.058	0.005
	(0.107)	(0.071)	(0.087)	(0.087)	(0.015)
Log income per capita around	0.419*	0.822***	0.555**	0.555**	0.904***
	(0.233)	(0.210)	(0.221)	(0.221)	(0.051)
Log population around	1.218***	1.063***	1.154***	1.154***	1.110***
	(0.105)	(0.069)	(0.085)	(0.085)	(0.010)
Log land area	0.162**	0.053	0.117*	0.117*	0.001
	(0.078)	(0.051)	(0.064)	(0.064)	(0.010)
Log income per capita	0.524**	0.175	0.405*	0.405*	-0.051
	(0.259)	(0.215)	(0.234)	(0.234)	(0.047)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.93	0.95	0.94	0.94	0.69
N	60,124	60,644	120,768	120,768	120,768
S.W. F stat: Log Population	11.34	10.49	12.25	12.25	22.7
S.W. F stat: Log Population × log frequency	22.49	18.8	22.68	22.68	15.36
S.W. F stat: Log Population × CEX expenditure share	14.82	16.7	16.4	16.4	16.45

Standard errors clustered at county level in parentheses.

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

Table E.5: **Local establishments around and frequency of purchase (2SLS)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	neighboring counties-sector log number of establishments				
Sample years :	07	98	98,07	98,07	98,07
Log population	-0.434***	-0.351***	-0.401***	-0.160***	-0.128***
	(0.134)	(0.112)	(0.123)	(0.052)	(0.026)
Log population × log frequency	0.074***	0.057***	0.066***	0.063***	0.064***
	(0.017)	(0.015)	(0.016)	(0.015)	(0.016)
Log population × CEX expenditure share	1.161***	1.289***	1.216***	1.209***	1.207***
	(0.197)	(0.199)	(0.192)	(0.185)	(0.186)
Log land area	0.160*	0.114	0.143*	0.010	0.004
	(0.092)	(0.078)	(0.085)	(0.028)	(0.007)
Log income per capita	0.704**	0.666**	0.707**	0.104	-0.007
	(0.308)	(0.325)	(0.315)	(0.122)	(0.030)
Log land area around	-0.164	-0.085	-0.130	0.107***	0.142***
	(0.127)	(0.106)	(0.117)	(0.034)	(0.013)
Log income per capita around	0.060	0.075	0.039	0.557***	0.602***
	(0.278)	(0.318)	(0.297)	(0.086)	(0.035)
Log population around	1.135***	1.059***	1.105***	0.891***	0.896***
	(0.124)	(0.105)	(0.114)	(0.036)	(0.008)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.91	0.93	0.92	0.92	0.68
N	60,124	60,644	120,768	120,768	120,768
S.W. F stat: Log Population	11.34	10.49	12.25	9.03	22.7
S.W. F stat: Log Population × log frequency	22.49	18.8	22.68	18.67	15.36
S.W. F stat: Log Population × CEX expenditure share	14.82	16.7	16.4	15.84	16.45

Standard errors clustered at county level in parentheses.

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

Table E.6: **Employees per establishments around and frequency of purchase (2SLS)**

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	neighboring counties-sector log number of employees per estab.				
Sample years :	07	98	98,07	98,07	98,07
Log population	0.187***	0.288***	0.232***	0.167***	0.166***
	(0.061)	(0.090)	(0.069)	(0.056)	(0.027)
Log population × log frequency	-0.154***	-0.135***	-0.145***	-0.147***	-0.151***
	(0.022)	(0.020)	(0.021)	(0.021)	(0.022)
Log population × CEX expenditure share	0.134	-0.551***	-0.194*	-0.227**	-0.253**
	(0.132)	(0.134)	(0.112)	(0.112)	(0.114)
Log land area around	0.045	0.111	0.072	-0.071**	-0.136***
	(0.055)	(0.085)	(0.064)	(0.035)	(0.013)
Log income per capita around	0.359***	0.747***	0.516***	0.270***	0.302***
	(0.119)	(0.257)	(0.164)	(0.093)	(0.032)
Log population around	0.082	0.004	0.049	0.159***	0.214***
	(0.053)	(0.084)	(0.063)	(0.039)	(0.007)
Log land area	0.002	-0.061	-0.026	0.013	-0.003
	(0.040)	(0.062)	(0.047)	(0.030)	(0.008)
Log income per capita	-0.180	-0.491*	-0.302*	-0.110	-0.044
	(0.132)	(0.263)	(0.175)	(0.133)	(0.037)
Sector Fixed Effects	Yes	Yes	Yes	No	No
Year Fixed Effects	No	No	Yes	No	No
Sector-Year Fixed Effects	No	No	No	Yes	Yes
State-Year Fixed Effects	No	No	No	Yes	No
Commuting Zone-Year Fixed Effects	No	No	No	No	Yes
R-square	0.22	0.11	0.18	0.20	-0.03
N	60,124	60,644	120,768	120,768	120,768
S.W. F stat: Log Population	11.34	10.49	12.25	9.03	22.7
S.W. F stat: Log Population × log frequency	22.49	18.8	22.68	18.67	15.36
S.W. F stat: Log Population × CEX expenditure share	14.82	16.7	16.4	15.84	16.45

Standard errors clustered at county level in parentheses.

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

E.3 Density and Sorting on Preferences

In this subsection, we examine the role of density and sorting on preferences.

In Table E.7 we extend our main regressions to control for an interaction between log income and log frequency of transactions. In this way, we control for the possibility that people with "tastes" for higher frequency sectors - as proxied by income - are more likely to sort in bigger places. In this table, the outcome variables are log employment (columns (1) and (2)), log establishment density (columns (3) and (4)), and log employees per establishment (columns (5) and (6)): within each of the outcomes, we replicate columns (5) and (6) in our main specification: columns differ on whether we control for heterogeneous state time trends or commuting zone time trends; all columns include sector-year fixed effects. As in our main specification, a higher income per capita is associated to a higher employment across sectors; moreover, employment tends to be smaller in high-frequency than in low-frequency sectors as income grows. The sign of this correlation may reflect non-homothetic preferences and possibly capture further un-modeled equilibrium feedbacks. We do not see this exercise as a way to identify the causal effect of higher income, but rather as a simple way to control for heterogeneity in individual characteristics. Importantly, the sign and magnitude of the interaction between population and frequency is little affected.

Table E.7: **Local outcome responses controlling for heterogeneity via income per capita**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log population	0.832***	1.236***	0.534***	0.790***	0.298***	0.446***
	(0.117)	(0.079)	(0.088)	(0.054)	(0.052)	(0.046)
Log population × log frequency	0.072**	0.063**	0.122***	0.118***	-0.050**	-0.055***
	(0.031)	(0.028)	(0.024)	(0.023)	(0.020)	(0.020)
Log population × CEX expenditure share	0.702*	0.622*	1.205***	1.143***	-0.503*	-0.521*
	(0.366)	(0.359)	(0.207)	(0.203)	(0.298)	(0.298)
Log land area	0.184***	-0.026	0.161***	0.033	0.023	-0.059**
	(0.034)	(0.044)	(0.026)	(0.029)	(0.015)	(0.026)
Log income per capita	2.610***	1.829***	2.263***	1.741***	0.348**	0.088
	(0.346)	(0.222)	(0.263)	(0.153)	(0.149)	(0.131)
Log income per capita × log frequency	-0.770***	-0.757***	-0.676***	-0.670***	-0.094*	-0.087
	(0.084)	(0.079)	(0.069)	(0.065)	(0.054)	(0.054)
Log income × CEX expenditure share	-1.837*	-1.594	-3.215***	-3.015***	1.378*	1.422*
	(1.003)	(0.978)	(0.582)	(0.563)	(0.814)	(0.814)
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	0.73	0.70	0.79	0.77	0.23	0.19
N	121,336	121,336	121,336	121,336	121,336	121,336
S.W. F stat: Log population	25.25	28.65	25.25	28.65	25.25	28.65
S.W. F stat: Log population × log frequency	24.54	23.6	24.54	23.6	24.54	23.6
S.W. F stat: Log population × CEX expenditure share	27.69	21.37	27.69	21.37	27.69	21.37

Standard errors clustered at county level in parentheses.

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

In Table E.8, we directly examine the role of density, again on log employment (columns (1) and (2)), log establishment density (columns (3) and (4)), and log employees per establishment (columns (5) and

(6)). The structure of the table mimics the one of Table E.7. The main regressors now include log income per capita, log density (computed as log population less log land area), an interaction of log density with log frequency, and the main object of our analysis, log population times log frequency. We now have five endogenous variables: log density, its interaction with log frequency and CEX expenditure share, and the interaction of population with frequency and CEX expenditure share: as instruments, we use the county composition of consolidated and semi-consolidated aquifers (first set of instrument), their interaction with frequency and CEX share (second and third set of instruments), and their interaction with county land size times frequency, or times CEX shares (fourth and fifth set of instruments). These regressions are essentially "fixing" a density level and compare, say, a small sprawling place to a large sprawling place. We find that, controlling for density, an increase in population is significantly associated to a larger employment increase in high-frequency sectors than in low-frequency sectors. The point estimate of the interaction effect is much larger. The reason can be traced to the interaction of density and frequency. We find that as we increase density, employment in high-frequency sectors *falls* relative to employment in low-frequency sectors. This is consistent with our results in Figure 5 in the main text, that shows that adding denser counties to the estimation sample tends to pull the interaction coefficient down and towards zero.

Table E.8: **Local outcome responses controlling for density**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log density	0.373***	0.668***	0.230***	0.417***	0.143***	0.250***
	(0.114)	(0.088)	(0.087)	(0.067)	(0.039)	(0.037)
Log density × log frequency	-0.272***	-0.268***	-0.171***	-0.162***	-0.101***	-0.107***
	(0.042)	(0.034)	(0.032)	(0.025)	(0.015)	(0.014)
Log density × CEX expenditure share	-1.347***	-1.352***	-0.955***	-0.931***	-0.392***	-0.421***
	(0.269)	(0.226)	(0.194)	(0.159)	(0.129)	(0.122)
Log population × log frequency	0.611***	0.582***	0.449***	0.415***	0.162***	0.167***
	(0.037)	(0.041)	(0.027)	(0.028)	(0.019)	(0.021)
Log population × CEX expenditure share	3.897***	3.757***	3.472***	3.289***	0.425	0.468
	(0.441)	(0.441)	(0.252)	(0.263)	(0.303)	(0.300)
Log income per capita	1.936***	1.608***	1.586***	1.395***	0.349***	0.213**
	(0.306)	(0.235)	(0.232)	(0.182)	(0.115)	(0.108)
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	0.65	0.60	0.70	0.66	0.20	0.17
N	121,336	121,336	121,336	121,336	121,336	121,336
S.W. F stat: Log density	27.84	34.79	27.84	34.79	27.84	34.79
S.W. F stat: Log density × log frequency	314.34	294.45	314.34	294.45	314.34	294.45
S.W. F stat: Log density × CEX expenditure share	231.95	224.27	231.95	224.27	231.95	224.27
S.W. F stat: Log population × log frequency	13.58	18.25	13.58	18.25	13.58	18.25
S.W. F stat: Log population × CEX expenditure share	16.57	20.02	16.57	20.02	16.57	20.02

Standard errors clustered at county level in parentheses.

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

E.4 Controlling for fixed costs

In this section we control for the interaction between a measure of fixed costs and frequency of transactions. A reasonable proxy is the economy-wide ratio of total employment to total establishments in a sector-year, i.e., the average establishment size: if fixed costs are high, increasing returns to scale tend to be important, and we should expect a higher employees-to-establishment ratio. In what follows, we will refer to this ratio simply as "fixed costs". We compute this ratio as the simple average of the employees per establishment in 1998 and 2007, our two sample years. Table E.9 replicates the most conservative specifications in columns (5) and (6) for Tables 6-8: all the unreported columns behave similarly. Columns (1) and (2) replicate Table 6, where the dependent variable is employment. The coefficient on the interaction between frequency and population stays positive and of very similar magnitude. Moreover, in response to larger population, sectoral employment does not seem to change differentially as a function of fixed costs. The remaining 4 columns look at the margins of these changes: columns (3) and (4) replicate columns (5) and (6) of Table 7, where the dependent variable is the log number of establishments; columns (5) and (6) replicate the last two columns of Table 8, which consider employees per store. We find that controlling for fixed costs tilt these interactions slightly towards zero; the interaction with fixed costs are not estimated to be significant per se. Overall, we read these results as further evidence that consumers' mobility impacts local economic outcomes.

Table E.9: **Local outcome responses controlling for fixed costs**

Dependent variable:	county-sector log employment		county-sector log establishments		county-sector log employees per estab.	
	(1)	(2)	(3)	(4)	(5)	(6)
Log population	0.709***	1.121***	0.493***	0.758***	0.216**	0.362***
	(0.128)	(0.118)	(0.131)	(0.097)	(0.084)	(0.100)
Log population × log frequency	0.091**	0.077**	0.110***	0.102***	-0.019	-0.026
	(0.041)	(0.038)	(0.031)	(0.029)	(0.023)	(0.023)
Log population × CEX expenditure share	0.765*	0.677	1.392***	1.321***	-0.627*	-0.644*
	(0.458)	(0.443)	(0.263)	(0.256)	(0.365)	(0.363)
Log population × log avg. employees per establishment	0.038	0.036	0.015	0.012	0.022	0.024
	(0.032)	(0.031)	(0.023)	(0.023)	(0.034)	(0.034)
Log land area	0.183***	-0.024	0.161***	0.036	0.023	-0.060**
	(0.034)	(0.043)	(0.026)	(0.028)	(0.015)	(0.026)
Log income per capita	1.766***	1.013***	1.444***	0.945***	0.323**	0.068
	(0.294)	(0.202)	(0.216)	(0.133)	(0.125)	(0.118)
Sector-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State-Year Fixed Effects	Yes	No	Yes	No	Yes	No
Commuting Zone-Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-square	0.73	0.69	0.78	0.76	0.23	0.20
N	121,336	121,336	121,336	121,336	121,336	121,336
S.W. F stat: Log population	19.07	24.57	19.07	24.57	19.07	24.57
S.W. F stat: Log population × log frequency	23.24	24.46	23.24	24.46	23.24	24.46
S.W. F stat: Log population × CEX expenditure share	21.41	22.62	21.41	22.62	21.41	22.62
S.W. F stat: Log population × log avg. employees per establishment	23.42	23.05	23.42	23.05	23.42	23.05

Standard errors clustered at county level in parentheses.

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